DESIGN OF PAIRED COMPARISON EXPERIMENTS
WITH QUANTITATIVE INDEPENDENT VARIABLES

BY

WALTER WILLIAM OFFEN

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1980

To Susie

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate Council
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DESIGN OF PAIRED COMPARISON EXPERIMENTS
WITH QUANTITATIVE INDEPENDENT VARIABLES

By

Walter William Offen

August 1980

Chairman: Ramon C. Littell
Major Department: Statistics

An experiment in which the treatments are compared and ranked
pairwise is called a paired comparison experiment. The Bradley-Terry
model for preference probabilities with ties not allowed is used in
this dissertation. This model defines treatment parameters, π_1, ... ,
π_t, such that the probability that treatment T_i is preferred over
treatment T_j is equal to $\pi_i(\pi_i+\pi_j)^{-1}$. When the treatments are levels
of continuous, quantitative variables, the logarithm of the Bradley-
Terry parameters can be modeled as a regression-type model,
$\ln\pi_i = \sum_k \beta_k x_{ki}$. There been a large number of papers on the topic
of paired comparisons, but only a few have considered experimental
design. Past papers on design are reviewed in Chapter 1. In 1973
Springall presented the asymptotic distribution of the maximum likeli-
hood estimators of β_1, ... ,β_s, where s is the number of regression-
type parameters in the model. Chapter 2 of the present dissertation
contains a derivation of this asymptotic distribution. The remainder

of the dissertation is a consideration of designs for situations where the treatments are levels of a single quantitative variable.

The asymptotic variance-covariance matrix is a function of the true parameter values. Hence, the optimal designs also depend on the parameters. In previous papers on design, the Bradley-Terry parameters were taken to be equal when evaluating the variances and covariances. The present dissertation considers the design problem for various parameter values.

Notice that the general design problem is complex. It involves deciding how many levels and which particular levels should be chosen. Furthermore, it must be determined how often to compare each pair of treatments. In some cases this complexity was resolved. In other cases additional restrictions had to be imposed before optimal designs could be found.

Chapter 3 considers designs which minimize the variance of a single parameter estimate, $\hat{\beta}_i$. The linear, quadratic, and cubic models are considered. The optimal design in the linear case turns out to be comparisons of only the smallest level in the experimental region with the largest, for most values of the linear parameter. The designs for the quadratic and cubic cases depend more heavily on the parameter values than for the linear case.

Chapter 4 presents optimal designs for fitting a quadratic model. The optimality criteria considered are D-optimality, the minimization of the average variance of $\ln \hat{\pi}_i$, and the minimization of the maximum value of $\ln \hat{\pi}_i$. These designs also depend heavily on the parameters, although some overall design recommendations are given in the last section.

The remaining chapters contain a brief discussion of related topics. Chapter 5 is a consideration of designs which protect against the bias present when the true model is cubic. Chapter 6 is a discussion of designs for preliminary test estimators. Chapter 7 is a consideration of two-stage sampling. The first stage is a design which results with a small variance of the quadratic parameter estimate. A test of hypothesis then determines whether the second stage design should be one which is optimal for fitting a linear or a quadratic model. A discussion of the best error rate to use for the test of hypothesis is included.

The appendices contain computer programs that find maximum likelihood estimates of the Bradley-Terry parameters and $\beta_1, \ldots, \beta_s$. These programs are written in the languages APL and Fortran.

CHAPTER 1

DESIGN OF PAIRED COMPARISON EXPERIMENTS:
A LITERATURE REVIEW

1.1 Introduction

An experiment in which the treatments are compared and ranked pairwise is called a paired comparison experiment. Such an experiment requires a number of subjects which are usually individuals, but could conceivably be a machine, for instance. Each subject is required to compare two samples and decide which of the two is preferable for the attribute under study. For example, each subject could be asked to state which of two samples tastes better. Paired comparison experiments are often conducted in areas such as food testing where it may be difficult for an individual to quantify his like or dislike of a sample, but he may be able to readily decide which of the two samples is preferred.

Suppose that an experimenter wants to determine which of four brands of coffee most consumers prefer. There are a number of experimental procedures available. One possibility is that each individual be required to taste and rank all four brands from least favorable to most favorable. However, sensory fatigue may be a problem because after tasting the second or third sample, it may be difficult to perceive any difference between the subsequent sample and the first sample tasted.

Another possible procedure is to have each individual assign to each brand a score on some numerical scale. Sensory fatigue may again be a problem with this experimental procedure since all four brands are tasted by each individual. Additionally, there may be a degree of arbitrariness to choosing a numerical score for each treatment.

A third possibility is to conduct a paired comparison experiment. The problems mentioned in the two preceding paragraphs are not present for paired comparison experiments. That is to say, paired comparison experiments have a minimal amount of individual guesswork involved. As pointed out in the first paragraph, the complete experiment involves a number of individuals, each one choosing out of two samples the one which is preferable. For the particular example with four brands of coffee, 24 subjects might be utilized, thereby resulting in 4 replicates of each of the 6 distinct pairs of treatments.

The treatments can have a factorial arrangement. However, factorial paired comparison experiments have often been analyzed ignoring the factorial arrangement of the treatments. For example, Larmond, Petrasovits, and Hill (1968) conducted a 2×3 factorial paired comparison experiment, but they did not test for interaction or main effects as normally is done with factorial experiments. Although Abelson and Bradley (1954) presented a theoretical development, only relatively recently have tractable methods been available for testing interaction and main effects in factorial paired comparison experiments.

In the case of quantitative treatments, an analogue of multiple linear regression analysis is appropriate. For example, instead of four distinctly different brands of coffee, there may only be one brand

of coffee with four different levels of an additive. The design of paired comparison experiments for fitting response surfaces with one independent variable is discussed in Chapters 3-7.

1.2 Bradley-Terry Model

Denote the t treatments as $T_1, T_2, \ldots, T_t$. Let $n_{ij} \geq 0$ be the number of times T_i and T_j are compared, $i<j$, $i,j=1, \ldots, t$. We write $T_i \rightarrow T_j$ for "T_i is selected over T_j", where selection is based on the particular attribute under study. A general model defines $\binom{t}{2}$ functionally independent parameters, $0 \leq \pi_{ij} \leq 1$, such that

$$P(T_i \rightarrow T_j) = \pi_{ij}, \quad \pi_{ij} + \pi_{ji} = 1, \quad i \neq j, \ i,j=1,\ldots,t. \qquad (1.2.1)$$

Each distinct pair of treatment comparisons is a set of independent Bernoulli trials. The entire experiment is a combination of the $\binom{t}{2}$ independent sets.

Bradley and Terry (1952) proposed a basic model for paired comparisons. A treatment parameter π_i is associated with each treatment T_i, $i=1, \ldots, t$. The Bradley-Terry model is then defined to be

$$P(T_i \rightarrow T_j) = \pi_{ij} = \pi_i/(\pi_i + \pi_j), \quad i \neq j, \ i,j=1,\ldots,t. \qquad (1.2.2)$$

The right-hand side of (1.2.2) is invariant under scale changes, so a single constraint is imposed to make the treatment parameters unique. Two popular choices are $\sum \pi_i = 1$, or $\sum \ln \pi_i = 0$. This constraint is replaced with another constraint in Chapters 2-7.

Another model for paired comparisons was presented by Thurstone (1927). He used the concept of a subjective continuum on which only order can be perceived. Let μ_i represent the location point on the continuum for treatment T_i, i=1, ... ,t. An individual receives a sensation X_i in response to treatment T_i, with X_i distributed Normal$(\mu_i,1)$. The probability that T_i is preferred over T_j is then

$$P(T_i \rightarrow T_j) = P(X_i > X_j) = \frac{1}{\sqrt{2\pi}} \int_{-(\mu_i-\mu_j)}^{\infty} e^{-y^2/2} \, dy \, , \quad \begin{array}{l} i \neq j \\ i,j=1,\ldots,t. \end{array} \qquad (1.2.3)$$

Bradley (1953) noted that replacement of the normal density function in (1.2.3) by the logistic density function yields

$$P(T_i \rightarrow T_j) = \int_{-(\ln\pi_i-\ln\pi_j)}^{\infty} \frac{1}{4} \, \text{sech}^2(y/2) \, dy = \pi_i/(\pi_i+\pi_j) \, , \quad \begin{array}{l} i \neq j \\ i,j=1,\ldots,t. \end{array} \qquad (1.2.4)$$

This is the Bradley-Terry model with $\mu_i=\ln\pi_i$, i=1, ... ,t.

Thompson and Singh (1967) gave a psychophysical consideration by assuming a treatment T stimulates unobservable signals U_1, ... ,U_m which are then transmitted to the brain by m sense receptors. The U_i are independent and identically distributed, and m is taken to be fixed and large. The two variables considered for the response X to treatment T are

     (i) X is the average of the signals U_1, ... ,U_m,

     (ii) X is the maximum of the signals U_1, ... ,U_m.

By the Central Limit Theorem, asymptotic normality results for X in (i), which results with Thurstone's model. The random variable in (ii) has one of three distributions depending on the distribution of the U_i. If U_i, for example, is distributed exponentially, then the asymptotic distribution of X is the extreme value distribution. In any case, however, the Bradley-Terry model in (1.2.4) results from (ii).

Rao and Kupper (1967) generalized the Bradley-Terry model to allow for ties. A tie parameter $\theta \geq 1$ is introduced, and the model becomes

$$P(T_i \to T_j) = \int_{-(\ln\pi_i - \ln\pi_j)+\eta}^{\infty} \frac{1}{4} \operatorname{sech}^2(y/2) \, dy = \pi_i/(\pi_i + \theta\pi_j) \,, \qquad \begin{array}{c} i \neq j \\ i,j=1,\ldots,t, \end{array}$$

$$P(T_i = T_j) = \int_{-(\ln\pi_i - \ln\pi_j)-\eta}^{-(\ln\pi_i - \ln\pi_j)+\eta} \frac{1}{4} \operatorname{sech}^2(y/2) \, dy \qquad (1.2.5)$$

$$\begin{array}{c} i \neq j \\ i,j=1,\ldots,t, \end{array}$$

$$= \frac{(\theta^2 - 1)\pi_i\pi_j}{(\pi_i + \theta\pi_j)(\theta\pi_i + \pi_j)} \,,$$

where $\eta = \ln\theta$ is the threshold of perception for the judges. That is to say, if it is hypothesized that the two samples generate a random variable d which measures the difference between the samples, and furthermore that $|d| < \eta$, then the judge will not be able to discern a difference between the two samples, and will thus declare a tie. Note that the Bradley-Terry model is a special case with $\theta=1$. Another model to handle ties was proposed by Davidson (1970).

These equations must be solved iteratively (see Appendix A for an APL computer program which finds maximum likelihood estimates of the treatment parameters). A brief description of the iterative procedure follows. Letting $p_i^{(k)}$ be the kth approximation to p_i, then

$$p_i^{(k)} = p_i^{*(k)} / \sum_{i=1}^{t} p_i^{*(k)} \quad , \tag{1.3.3}$$

where

$$p_i^{*(k)} = a_i / \sum_{\substack{j \\ j \neq i}}^{t} \left[n_{ij}/(p_i^{(k-1)} + p_j^{(k-1)}) \right] \quad , \quad k=1,2,\cdots. \tag{1.3.4}$$

The iteration can be started with $p_i^{(0)} = 1/t$, $i=1, \ldots ,t$. Dykstra (1956) suggested a method of obtaining good initial parameter estimates $p_i^{(0)}$, $i=1, \ldots ,t$.

Ford (1957) proposed the model independently and proved that the iterative procedure converged to a unique maximum for L if the following assumption is true.

Assumption 1.1. For every possible partition of the treatments into the two sets S_1 and S_2, some treatment in S_1 is preferred at least once to some treatment in S_2.

If Assumption 1.1 does not hold, then a_i must be zero for at least one i. If exactly one a_i is zero, say a_{i_o}, there may still be a unique solution with $p_{i_o} = 0$. However, Ford proceeded to explain that if the set S_2 which violates Assumption 1.1 contains two or more treatments, then the corresponding estimates of the Bradley-Terry treatment

parameter must be zero, and consequently individual members of S_2 could not be compared. This fact is also evident by noticing that in (1.3.4), $p_i^{*(k)}$ would be zero for at least two values of i, and so by (1.3.3), $p_i^{(k)}$ would also be zero for the same values of i. Then the next time in the iterative procedure that (1.3.4) is executed, the denominator in the summation would be zero for some i and j, and hence the iterative procedure would not converge.

The asymptotic theory and tests of hypothesis require the following additional assumption.

Assumption 1.2. For every partition of the treatments into two non-empty subsets S_1 and S_2, there exists a treatment $T_i \varepsilon S_1$ and a treatment $T_j \varepsilon S_2$ such that $\mu_{ij} > 0$, where

$$\mu_{ij} = \lim_{N \to \infty} \frac{n_{ij}}{N}, \qquad i \neq j, \ i,j = 1, \ldots, t,$$

$$N = \sum_{i<j} \sum n_{ij}.$$

Bradley (1955) investigated asymptotic properties of the treatment parameters with each $n_{ij} = n$. Davidson and Bradley (1970) considered a multivariate model where each subject chooses one of two samples on more than one attribute. As a special case of the multivariate model, Davidson and Bradley obtained a more general result, allowing the n_{ij} to be different. They showed that the asymptotic joint distribution of $\sqrt{N}(p_1 - \pi_1), \ldots, \sqrt{N}(p_t - \pi_t)$ is the singular multivariate normal distribution of dimensionality (t-1) in a space of t dimensions, with zero mean

vector and variance-covariance matrix $\Sigma = (\sigma_{ij})$, where

$$
\sigma_{ij} = \text{cofactor of } \lambda_{ij} \text{ in } \cfrac{\begin{bmatrix} (\lambda_{ij}) \ _{t\times t} & \underline{1} \ _{t\times 1} \\[2em] \underline{1}' \ _{1\times t} & 0 \end{bmatrix}}{\begin{vmatrix} (\lambda_{ij}) \ _{t\times t} & \underline{1} \ _{t\times 1} \\[2em] \underline{1}' \ _{1\times t} & 0 \end{vmatrix}} \quad , \tag{1.3.5}
$$

$\underline{1}' = (1,1, \ldots ,1)$, the t-dimensional unit row vector, and

$$
\lambda_{ii} = \frac{1}{\pi_i} \sum_{\substack{j \\ j\neq i}}^{t} \mu_{ij} \frac{\pi_j}{(\pi_i + \pi_j)^2} \quad , \qquad\qquad i=1,\ldots,t,
$$

$$
\lambda_{ij} = -\mu_{ij}/(\pi_i + \pi_j)^2 \quad , \qquad\qquad i\neq j,\ i,j=1,\ldots,t.
$$

It is apparent that the variances and covariances in (1.3.5) depend on the true treatment parameters. Hence estimates of the variances and covariances are usually found by substituting the consistent maximum likelihood estimates for $\pi_1, \ldots ,\pi_t$ into (1.3.5).

In the special case with $\pi_i=1/t$, $n_{ij}=n$ for all i,j,

$$
\sigma_{ii} = 2(t-1)^2/t^3 \quad , \quad \text{and} \quad \sigma_{ij} = -2(t-1)/t^3 \quad , \quad i\neq j. \tag{1.3.6}
$$

Use of (1.3.6) is adequate if $\pi_1, \ldots ,\pi_t$ are approximately equal.

The major test proposed is a test of treatment equality,

$$H_o: \quad \pi_1 = \pi_2 = \cdots = \pi_t = 1/t \ ,$$

versus

$$H_a: \quad \pi_i \neq \pi_j \ , \quad \text{for some } i,j, \ i \neq j, \ i,j=1,\ldots,t.$$

The likelihood ratio statistic is

$$-2\ln\lambda \ = \ 2N\ln 2 \ - \ 2 \left[\sum_{i<j} \sum n_{ij}\ln(p_i+p_j) \ - \ \sum_i a_i\ln p_i \right] \ .$$

Under the null hypothesis, H_o, $-2\ln\lambda$ is asymptotically central chi-square with $(t-1)$ degrees of freedom. Bradley and Terry (1952) and Bradley (1954) provided tables for $t=3,4,5$, including exact significance levels. Comparison of the chi-square significance levels with the tabled exact significance levels suggests that the former may be used even for relatively small values of n_{ij}. This is perhaps comparable to using the normal approximation to the binomial.

The asymptotic distribution under the alternative hypothesis is also available for $-2\ln\lambda$. Let

$$\pi_i \ = \ \frac{1}{t} + \frac{\delta_{iN}}{\sqrt{N}} \ , \qquad \sum_{i=1}^{t} \delta_{iN} \ = \ 0 \ , \qquad \lim_{N\to\infty} \delta_{iN} \ = \ \delta_i \ .$$

Then $-2\ln\lambda$ is asymptotically noncentral chi-square with $(t-1)$ degrees of freedom and noncentrality parameter λ_1, where

$$\lambda_1 \ = \ \frac{t^2}{4} \sum_{i<j} \sum \mu_{ij}(\delta_i - \delta_j)^2 \ .$$

1.4 Designs of Factorial Paired Comparison Experiments

Littell and Boyett (1977) compared two designs for R×C factorial paired comparison experiments. The two designs considered are:

Design I: Consists of independent paired comparisons of each of the $\binom{RC}{2}$ pairs of treatments.

Design II: Consists of comparisons between pairs of treatment combinations differing only in the levels of one factor, i.e. comparing two treatments in the same row or column.

Note that there are $\binom{RC}{2} = \frac{1}{2}RC(RC-1)$ distinct pairs of treatments for Design I, but only $R\binom{C}{2} + C\binom{R}{2} = \frac{1}{2}RC(R+C-2)$ distinct pairs of treatments for Design II. In the 4×4 case, for example, Design I requires 120 distinct paired comparisons whereas Design II requires only 48 distinct paired comparisons. The two designs could entail the same total number of paired comparisons, but Design II requires preparation of fewer distinct pairs of samples.

Design II is also preferable from a computational point of view because its analysis can be carried out by using a computer program capable of handling the one-way classification with unequal replication of pairs. A specialized factorial program is required for Design I to obtain maximum likelihood estimates under the main effects hypothesis, H_o: $\pi_{ij} = \alpha_i \beta_j$, where π_{ij} is the Bradley-Terry treatment parameter given by (1.4.1).

The two designs are illustrated for the 2×2 case in Figure 1.1. The comparisons made with Design I are indicated by the solid arrows. The comparisons made with Design II are indicated by the dotted arrows.
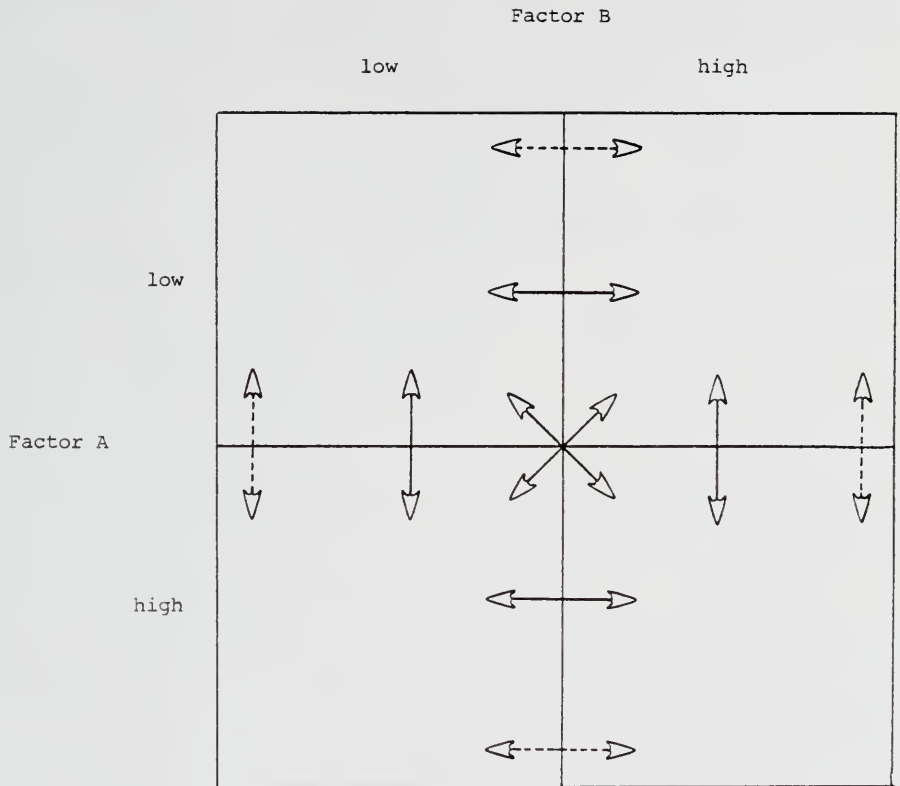
Designs I and II were compared by Littell and Boyett on the basis of the asymptotic relative efficiencies, in the Bahadur sense, of their respective likelihood ratio tests. That is, test statistics are compared on the basis of the ratio of the exponential rates of convergence to zero of their observed significance levels. In general, the Bahadur efficiency of test 1 relative to test 2 is defined to be

$$E = \frac{\lim_{n\to\infty}(-\frac{2}{n}\log L_n^{(1)})}{\lim_{n\to\infty}(-\frac{2}{n}\log L_n^{(2)})} = c_1(\theta)/c_2(\theta) \ ,$$

where $L_n^{(i)}$ is the observed significance level of test i. The function $c_i(\theta)$ is called the exact slope of test i.

Usually Bahadur efficiency is used to compare two test statistics that are based on the same sample space. However, its use is equally appropriate when the sample spaces are different. The two designs should be compared on the basis of an equivalent number of comparisons in each. Therefore, a "single observation" with Design I will be taken as n_I = R+C-2 comparisons of each pair in the design, and a "single observation" with Design II will be taken as n_{II} = RC-1 comparisons of each pair in the design. This gives $\frac{1}{2}$RC(RC-1)(R+C-2) total comparisons in each design.

Let T_{ij} represent the treatment with the first factor at level i and the second factor at level j. The Bradley-Terry model is then

$$P(T_{ij} \to T_{kl}) = \pi_{ij}/(\pi_{ij} + \pi_{kl}) \ , \qquad \begin{array}{l} i,k=1,\ldots,R, \\ j,l=1,\ldots,C. \end{array} \qquad (1.4.1)$$

The following tests of hypotheses were considered by Littell and Boyett.

Test 1: H_o: $\pi_{ij} = 1/RC$, H_a: $\pi_{ij} = \alpha_i \beta_j$.
This is a test for detecting two main effects versus no treatment differences. Efficiencies were tabulated for the 2×2 case (Littell and Boyett (1977), Table 1). The table showed that Design I is more efficient than Design II for most combinations of α and β.

Test 2: H_o: $\pi_{ij} = 1/RC$, H_a: $\pi_{ij} = \beta_j/R$.
This is a test for detecting a single main effect versus no treatment differences. For this test, Design I was shown to be uniformly more efficient than Design II.

Test 3: H_o: $\pi_{ij} = \beta_j/R$, H_a: $\pi_{ij} = \alpha_i \beta_j$.
This is a test for detecting two main effects versus a single main effect. Efficiencies were again tabulated for the 2×2 case. The table showed that Design I is more efficient than Design II for most combinations of α and β.

Test 4: H_o: $\pi_{ij} = \alpha_i \beta_j$, H_a: $\pi_{ij} \geq 0$.
This is a test for detecting interaction versus two main effects. The results showed that Design II is uniformly superior to Design I for detecting interaction.

Summarizing, when testing for main effects (Tests 1-3), Design I is usually more efficient than II. When testing for interaction, Design II is uniformly more efficient than I.

Littell and Boyett suggested that it might be a good idea for the experiment to be conducted in two stages. The first stage is a number of replicates of Design II followed by a test for interaction. If it is significant, the second stage is completed by again using Design II, and simple effects are analyzed by applying one-way analyses within rows and columns. If interaction is not significant, then the second stage consists of replicates of pairs in Design I that are not in Design II, and main effects are analyzed.

The total number of comparisons to be made in the experiment can be divided into the two parts such that after the experiment is completed, the end result is either a complete Design I or Design II. For example, suppose the treatments form a 2×2 factorial, and resources are available for 24 comparisons. In the first stage, the experimenter runs 4 replicates of the four comparisons in Design II (see Figure 1.1). If interaction is significant, two more replicates of the four comparisons in Design I that are not in Design II are run. The latter then results in the same comparisons made as if a single stage experiment using Design I had been conducted.

Quenouille and John (1971) also considered designs for 2^n factorials in paired comparisons. However, they assumed an individual is asked to assign a number on a scale to signify the degree of difference in the two samples instead of merely deciding which sample possesses more of the particular attribute under study. This model was developed by Scheffe (1952).

In particular, Quenouille and John discussed the following 2^3 factorial. The 28 possible paired comparisons were divided into 7 sets

of 4 blocks as follows:

Set (1):	(I,ABC)	(A,BC)	(B,AC)	(C,AB)
Set (2):	(I,AB)	(A,B)	(C,ABC)	(AC,BC)
Set (3):	(I,AC)	(A,C)	(B,ABC)	(AB,BC)
Set (4):	(I,BC)	(A,ABC)	(B,C)	(AB,AC)
Set (5):	(I,A)	(B,AB)	(C,AC)	(BC,ABC)
Set (6):	(I,B)	(A,AB)	(C,BC)	(AC,ABC)
Set (7):	(I,C)	(A,AC)	(B,BC)	(AB,ABC) .

(1.4.2)

The letter "I" represents the treatment where all factors are at their low levels. If a letter appears, then that factor is at its high level. Note that each treatment appears once in each set. These sets are constructed by first pairing treatment I with each of the remaining treatments. These are called the initial blocks. The remaining blocks in the sets are constructed by multiplying the initial block by any treatment that has not yet appeared in that particular set, with the usual rule that $A^2 = B^2 = C^2 = 1$.

To illustrate this technique, we will construct set (1) in (1.4.2). The initial block is (I,ABC). So far only I and ABC have appeared, so we may choose any treatment other than these. Suppose we choose BC. Then the next block in the set would be (BC,A). Note that order is unimportant. Now we may choose any treatment other than I, ABC, BC, or A. If we choose B, we get the third block in set (1). To construct the fourth block, there are only two choices remaining, namely C or AB. Either one results in the last block of set (1).

Each set lacks information on the effects that have an even number of letters in common with the initial block. For example, the initial

block for set (1) is (I,ABC), so this set provides no information on

the effects AB, AC, and BC, i.e. on all first-order interactions. Sim-

ilarly, the initial block for set (5) in (1.4.2) is (I,A), so this set

provides no information on B, C, and BC.

This concept of no information on B, C, and BC from data in set

(5) may be viewed as follows. If the responses from comparisons in set

(5) are modeled as

$$Y_{(I,A)} = (\mu+a+e_A) - (\mu+e_I) = a+\varepsilon_{(I,A)} \; ,$$

$$Y_{(B,AB)} = (\mu+a+b+(ab)+e_{AB}) - (\mu+b+e_B) = a+(ab)+\varepsilon_{(B,AB)} \; ,$$

$$Y_{(C,AC)} = (\mu+a+c+(ac)+e_{AC}) - (\mu+c+e_C) = a+(ac)+\varepsilon_{(C,AC)} \; ,$$

$$Y_{(BC,ABC)} = (\mu+a+b+c+(ab)+(ac)+(bc)+(abc)+e_{ABC}) -$$

$$(\mu+b+c+(bc)+e_{BC})$$

$$= a+(ab)+(ac)+(abc)+\varepsilon_{(BC,ABC)} \; ,$$

where all of the e's are independent and have expected value zero, then

it is clear that b, c, and (bc), corresponding to the effects B, C, and

BC, are not estimable from the four comparisons in set (5).

Quenouille and John define efficiency as follows. Suppose the

design is made up of r sets, s (s $\leq$ r) of which give information on some

particular effect. Then s/r is the fraction of the comparisons made

that give information on that particular effect. On the other hand,

suppose that each set is run exactly once, i.e. each pair of treatments

is compared. It can be shown that there are a total of (2^n-1) sets,

2^{n-1} of which give information on any particular effect. In this case,

$2^{n-1}/(2^n-1)$ is the fraction of the comparisons made that give

Table 1.1.

Efficiencies of balanced 2^2 and 2^3 factorial paired comparison designs

n	Design	Number of blocks	Efficiency of interactions of order		
			0	1	2
2	{2}	2	1.50	0.00	
	{1}	4	0.75	1.50	
3	{3}	4	1.75	0.00	1.75
	{2}	12	1.17	1.17	0.00
	{1}	12	0.58	1.17	1.75
	{3,2}	16	1.31	0.88	0.44
	{3,1}	16	0.88	0.88	1.75
	{2,1}	24	0.88	1.17	0.88

Notation: {j}: A design constructed by combining all sets with initial block I and a j-factor combination.

{j,k}: A design constructed by combining the designs {j} and {k}.

Source: Quenouille and John (1971, Table 1).

information on any particular effect. The efficiency defined by Quenouille and John is the ratio of these two fractions. In other words, the efficiency is defined to be

$$E = \frac{s/r}{2^{n-1}/(2^n-1)} \quad .$$

Table 1.1 gives the efficiencies of some designs in the 2^2 and 2^3 cases. These designs are balanced in the sense that all main effects are estimated with the same efficiency, all first order interactions

are estimated with the same efficiency, and so on. The table indicates
that in the 2^2 case, Design {1} is efficient for estimating the inter-
action effect. This result is the same as the result Littell and
Boyett obtained using Bahadur efficiency.

El-Helbawy and Bradley (1978) considered treatment contrasts in
paired comparison experiments. The theory developed is completely gen-
eral, but they specifically showed how optimal designs can be found in
the case of a 2^3 factorial. Their optimality criterion is the maximi-
zation of the chi-square noncentrality parameter from the likelihood
ratio test, which thereby maximizes the asymptotic power of the test.

Let $\underline{a}=(a_1,a_2,a_3)$, where a_i represents the level of factor i,
i=1,2,3, and each a_i is 0 or 1. Then the treatment parameters are
defined by El-Helbawy and Bradley to be

$$\pi_{\underline{a}} = \pi_{a_1}^1 \pi_{a_2}^2 \pi_{a_3}^3 \pi_{a_1 a_2}^{12} \pi_{a_1 a_3}^{13} \pi_{a_2 a_3}^{23} \pi_{a_1 a_2 a_3}^{123} , \qquad (1.4.3)$$

for all $\underline{a}$. The factorial parameters are defined by taking logarithms
of the 8 treatment parameters given by (1.4.3). So from (1.4.3),

$$\gamma_{\underline{a}} = \log \pi_{\underline{a}}$$

$$= \sum_{i=1}^{3} \log \pi_{a_i}^i + \sum \sum_{i<j} \log \pi_{a_i a_j}^{ij} + \log \pi_{a_1 a_2 a_3}^{123}$$

$$= \sum_{i=1}^{3} \gamma_{a_i}^i + \sum \sum_{i<j} \gamma_{a_i a_j}^{ij} + \gamma_{a_1 a_2 a_3}^{123} . \qquad (1.4.4)$$

The last two expressions in (1.4.4) define the factorial parameters

$\gamma_{a_1}^{1}$, $\gamma_{a_2}^{2}$, $\gamma_{a_3}^{3}$, $\gamma_{a_1 a_2}^{12}$, $\gamma_{a_1 a_3}^{13}$, $\gamma_{a_2 a_3}^{23}$, $\gamma_{a_1 a_2 a_3}^{123}$. These parameters are subject to the usual analysis of variance constraints, namely

$$\sum_{a_i=0}^{1} \gamma_{a_i}^{i} = 0 , \qquad\qquad i=1,2,3,$$

$$\sum_{a_i=0}^{1} \gamma_{a_i a_j}^{ij} = \sum_{a_j=0}^{1} \gamma_{a_i a_j}^{ij} = 0 , \qquad i<j,\ i,j=1,2,3, \qquad (1.4.5)$$

$$\sum_{a_1=0}^{1} \gamma_{a_1 a_2 a_3}^{123} = \sum_{a_2=0}^{1} \gamma_{a_1 a_2 a_3}^{123} = \sum_{a_3=0}^{1} \gamma_{a_1 a_2 a_3}^{123} = 0 .$$

The null hypothesis of no two-factor interaction between the first two factors is now considered. This test may be made by testing

$$H_o: \ \gamma_{00}^{12} = 0 .$$

The constraints then require that $\gamma_{a_1 a_2}^{12}=0$, $a_1, a_2 = 0,1$. The null hypothesis is therefore equivalent to

$$H_o: \ \gamma_{00}^{12} - \gamma_{01}^{12} - \gamma_{10}^{12} + \gamma_{11}^{12} = 0 ,$$

since by (1.4.5), $\gamma_{00}^{12} - \gamma_{01}^{12} - \gamma_{10}^{12} + \gamma_{11}^{12} = 4\gamma_{00}^{12}$. Let $\underline{B}_n$ denote a matrix whose rows are orthonormal contrasts. If we let $\underline{\gamma}' = (\gamma_{000}, \gamma_{001}, \gamma_{010}, \gamma_{011}, \gamma_{100}, \gamma_{101}, \gamma_{110}, \gamma_{111})$, then the null hypothesis is also equivalent to

$$H_o: \ \underline{B}_{n=1} \underline{\gamma} = 0 ,$$

where $\underline{B}_{n=1} = \frac{1}{8} (1,1,-1,-1,-1,-1,1,1)$.

Let

$$
\frac{n_{ij}}{N} =
\begin{cases}
a & \text{if comparing } T_i \text{ and } T_j \text{ give no information (in the} \\
& \text{sense discussed by Quenouille and John) on the two-} \\
& \text{factor interaction between the first two factors,} \\
\\
b & \text{otherwise,}
\end{cases}
$$

where n_{ij} is defined in Section 1.2, and $N = \displaystyle\sum_{i<j}\sum n_{ij}$.

Then $12a + 16b = 1$, since 16 of the 28 distinct pairs give information

on the two-factor interaction. El-Helbawy and Bradley then proceeded

to show that b should be taken as large as possible in order to maximize

the noncentrality parameter. This gives the result $b=1/16$, $a=0$. So

the optimal design is to compare the treatments that give information

on the two-factor interaction between the first two factors. These

pairs of treatments would be the sets (3), (4), (5), and (6) in (1.4.2).

This result is the same as Quenouille and John would obtain if

they wanted to find an optimal design to estimate AB interaction, since

using these 4 sets maximizes the efficiency of the AB interaction using

their definition of efficiency. It is also analogous to the result

Littell and Boyett obtained because the treatments compared in either

case differ either in the level of A, or the level of B, but not both.

El-Helbawy and Bradley noted that the result discussed in the two

preceding paragraphs still holds if other specified factorial effects

are assumed to be null (e.g. three-factor interaction). While the sta-

tistics for the two tests may differ, their limit distributions are the

same for either H_o or H_a.

El-Helbawy and Bradley also considered simultaneously testing for the presence of AB, AC, and ABC interactions. In this case the null hypothesis is

$$H_o: \underline{B}_{n=3}\underline{\gamma} = 0 ,$$

where

$$\underline{B}_{n=3} = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix} .$$

By the same method they concluded that the optimal design is to make the following comparisons (using Quenouille and John's notation): (I,A), (B,AB), (C,AC), and (BC,ABC). Note that this is set (5) in (1.4.2). Note also that this is the only set that contains information on all three of the effects AB, AC, and ABC. So this design is also optimal using Quenouille and John's definition of efficiency.

Beaver (1977) considered the analysis of factorial paired comparison experiments using the weighted least squares approach of Grizzle, Starmer, and Koch (1969). He considered a loglinear transformation on the π_i's. This approach produces noniterative estimates of the π_i's. Also, the usual tests for a factorial experiment are easily implemented. However, Beaver gave no consideration to experimental design.

1.5 Response Surface Designs

Springall (1973) considered response surface fitting using the Bradley-Terry model with ties allowed as generalized by Rao and Kupper (1967). The model is given in (1.2.5).

The logarithm of the treatment parameters are defined to be functions of continuous, independent variables, $x_1, x_2, \ldots, x_s$, measurable without error, i.e.

$$\ln \pi_i = \sum_{k=1}^{s} \beta_k x_{ki} , \qquad i=1,\ldots,t. \qquad (1.5.1)$$

Note that some of the variables can be functions of other variables. For example, s=2 and $x_{2i}=x_{1i}^2$ results in the parameterization $\ln \pi_i = \beta_1 x_{1i} + \beta_2 x_{1i}^2$. The parameters $\theta, \beta_1, \ldots, \beta_s$ are estimated by the maximum likelihood method. Appendix B has an APL program and a Fortran program which produce maximum likelihood estimates of $\beta_1, \ldots, \beta_s$, but which assume there are no ties.

Springall presented the following definition, theorem, and example.

<u>Definition 1.1.</u> Analogue designs are defined to be designs in which the elements of the paired comparison variance-covariance matrix are proportional to the elements of any classical response surface variance-covariance matrix with the same design points and an equal number of replicates at each design point.

The advantage of such designs is that they enable certain desirable properties found in classical response surface designs (e.g. rotatability, orthogonality), which are dependent on the relative sizes of the elements of the variance-covariance matrix, to be readily produced.

For example, consider the classical regression model $y = \beta_0^* + \beta_1^* x + \beta_2^* x^2$. Then an experiment can be designed in such a way that the variance-covariance matrix of the least squares estimators $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, and

$\hat{\beta}_2^*$ possesses an optimal form. For instance, the classical design with an equal number of replicates at each of the four levels x=-2,-1,1,2 can be shown to be rotatable, and so the analogue paired comparison designs discussed in Example 1.1 below have the same property, i.e. $\text{var}(\ln \hat{\pi}_x) = \text{var}(\ln \hat{\pi}_{-x})$.

The analagous paired comparison model is $\ln \pi_x = \beta_1 x + \beta_2 x^2$. An analogue paired comparison design has a corresponding variance-covariance matrix of the maximum likelihood estimators $\hat{\beta}_1, \hat{\beta}_2$, such that each element is a constant times the corresponding element of the least squares variance-covariance matrix. For the particular example of a quadratic model currently under discussion, $\text{var}(\hat{\beta}_1) = c \cdot \text{var}(\hat{\beta}_1^*)$, $\text{var}(\hat{\beta}_2) = c \cdot \text{var}(\hat{\beta}_2^*)$, and $\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = c \cdot \text{cov}(\hat{\beta}_1^*, \hat{\beta}_2^*)$, for some c.

Springall described a method which uses linear programming techniques that produces an analogue design with minimum elements of the variance-covariance matrix. The following theorem shows how one can produce an analogue design, but it does not necessarily have minimum elements of the variance-covariance matrix.

Theorem 1.1. If a classical response surface design has certain properties dependent on the relative sizes of the elements of the variance-covariance matrix, then the analogue paired comparison design will have the same properties if

$$n_{ij} = N \left[\phi_{ij} \sum \sum_{k<l} \phi_{kl}^{-1} \right]^{-1} ,$$

where $\phi_{ij} = \pi_i \pi_j / (\pi_i + \pi_j)^2$.

Example 1.1. Consider the quadratic model $\ln \pi_x = \beta_1 x + \beta_2 x^2$, with design coordinates x=-2,-1,1,2. Assume $\theta=1$ and $\beta_1=\beta_2=0$, and set N=150. Springall showed that an analogue design is given by $n_{ij} = 150/6 = 25$, i<j, i,j=1,2,3,4.

An alternative design may be obtained by using linear programming. Let Σ denote the variance-covariance matrix of the maximum likelihood estimates $\hat{\beta}_1$ and $\hat{\beta}_2$. By simultaneously maximizing the elements of Σ^{-1} subject to the constraint N=150, a design which minimizes the elements of Σ is produced. Springall did this and arrived at the design $\{n_{12}=0, n_{13}=71, n_{14}=9, n_{23}=0, n_{24}=70, n_{34}=0\}$, after integerization of the n_{ij}.

The asymptotic distribution of the maximum likelihood estimates of $\theta,\beta_1, \ldots ,\beta_s$ was presented by Springall, but he presented few details. Therefore, the maximum likelihood estimates of the parameters $\beta_1, \ldots ,\beta_s$ and their asymptotic joint distribution are developed in Chapter 2. The tie parameter, θ, is not considered in the remainder of this dissertation.

CHAPTER 2

ASYMPTOTIC THEORY OF THE
MAXIMUM LIKELIHOOD ESTIMATORS

2.1   Introduction

This chapter contains the asymptotic theory for the maximum like-
lihood estimates of the parameters given by (1.5.1), where the logarithms
of the treatment parameters are expressed as a polynomial function of a
single quantitative variable, x. An example of an experiment which had
a single quantitative variable was given in Section 1.1. The treatments
in that example were different levels of an additive for coffee. For
an experiment with quantitative independent variables, it is appropriate
to do a paired comparison analogue to classical regression analysis.

The theory developed in Sections 2.2 and 2.3 assume the general
case of s independent variables $x_1, \ldots, x_s$. The parameterization of
the Bradley-Terry parameters, given in (1.5.1), is reproduced here:

$$\ln \pi_i \;=\; \sum_{k=1}^{s} \beta_k x_{ki} \;, \qquad i=1,\ldots,t.$$

The Bradley-Terry model with ties not allowed is used, where the π_i are
defined by (1.2.2).

In order for all of the parameters $\beta_1, \ldots, \beta_s$ to be estimable, it
is required that $s \leq (t-1)$. This is clearly necessary since the Bradley-
Terry parameters have dimension (t-1). Note also that an intercept

26

parameter, β_0, is not estimable without imposing a constraint on the
Bradley-Terry parameters. This becomes apparent when examining the
preference probabilities. For example, suppose the model was

$$\ln \pi_i = \beta_0 + \sum_{k=1}^{s} \beta_k x_{ki} , \qquad i=1,\ldots,t.$$

Then

$$P(T_i \to T_j) = \pi_i / (\pi_i + \pi_j)$$

$$= \frac{\exp(\beta_0 + \sum_{k=1}^{s} \beta_k x_{ki})}{\exp(\beta_0 + \sum_{k=1}^{s} \beta_k x_{ki}) + \exp(\beta_0 + \sum_{k=1}^{s} \beta_k x_{kj})}$$

$$= \frac{\exp(\sum_{k=1}^{s} \beta_k x_{ki})}{\exp(\sum_{k=1}^{s} \beta_k x_{ki}) + \exp(\sum_{k=1}^{s} \beta_k x_{kj})} , \qquad \begin{array}{l} i \neq j \\ i,j=1,\ldots,t. \end{array}$$

The parameter β_0 falls out, implying that any choice of β_0 results with
the same preference probabilities. Choosing $\beta_0=0$ replaces the imposi-
tion of a constraint on the treatment parameters.

2.2 Estimation

As was previously mentioned, the parameters $\beta_1, \ldots, \beta_s$ are esti-
mated by the maximum likelihood method. From (1.3.1), the log-likeli-
hood equation is

$$\ln L = \sum_{i=1}^{t} a_i \ln \pi_i - \sum_{i<j} \sum n_{ij} \ln(\pi_i + \pi_j)$$

$$= \sum_{i=1}^{t} a_i \left[\sum_{k=1}^{s} \beta_k x_{ki} \right] -$$

$$\sum_{i<j} \sum n_{ij} \ln \left[\exp\left(\sum_{k=1}^{s} \beta_k x_{ki} \right) + \exp\left(\sum_{k=1}^{s} \beta_k x_{kj} \right) \right] .$$

(2.2.1)

The maximum likelihood estimators are found by taking partial derivatives of the log-likelihood equation, setting them equal to zero, and solving for the parameters. The likelihood equations are

$$\frac{\partial \ln L}{\partial \beta_r} = \sum_{i=1}^{t} a_i x_{ri} - \sum_{i<j} \sum n_{ij} \frac{x_{ri} \exp\left(\sum_{k=1}^{s} \beta_k x_{ki} \right) + x_{rj} \exp\left(\sum_{k=1}^{s} \beta_k x_{kj} \right)}{\exp\left(\sum_{k=1}^{s} \beta_k x_{ki} \right) + \exp\left(\sum_{k=1}^{s} \beta_k x_{kj} \right)}$$

$$= \sum_{i \neq j} \sum m_{ij} x_{ri} - \sum_{i \neq j} \sum m_{ij} \frac{x_{ri} \exp\left(\sum_{k=1}^{s} \beta_k x_{ki} \right) + x_{rj} \exp\left(\sum_{k=1}^{s} \beta_k x_{kj} \right)}{\exp\left(\sum_{k=1}^{s} \beta_k x_{ki} \right) + \exp\left(\sum_{k=1}^{s} \beta_k x_{kj} \right)}$$

$$= \sum_{i \neq j} \sum m_{ij} \pi_j \frac{x_{ri} - x_{rj}}{\pi_i + \pi_j} = 0 , \quad r=1,\ldots,s.$$

(2.2.2)

where m_{ij} is the total number of times T_i is preferred over T_j.

These equations must be solved iteratively. Appendix B contains
an APL program and a Fortran program which find the estimates of $\beta_1,\dots,$
β_s. The procedure described in the appendix converged for most situa-
tions. However, occasionally when an attempt was made to fit a full
model (s=t-1), the procedure did not converge, which remains unexplained.
In this situation one can resort to the program found in Appendix A
which finds estimates of the Bradley-Terry parameters, and then solve
(t-1) equations in (t-1) unknowns to find the estimates of $\beta_1, \dots ,\beta_s$.

2.3 Asymptotic Distribution

First of all, some additional notation and elementary theory is
presented. This will become useful when the asymptotic joint distribu-
tion of $\hat{\beta}_1, \dots ,\hat{\beta}_s$ is derived.

Let X_{ijk} be a random variable which is equal to 1 if T_i is selected
over T_j for the kth comparison of T_i and T_j, 0 otherwise. Then X_{ijk}
has the Bernoulli distribution with probability of "success" equal to
$\pi_i/(\pi_i + \pi_j)$, i,j=1, $\dots$,t, k=1, $\dots$,n_{ij}, and the Bernoulli random
variables are independent of each other. This implies that m_{ij} has the
binomial distribution, with sample size n_{ij} and "success" probability
equal to $\pi_i/(\pi_i + \pi_j)$, where

$$m_{ij} = \sum_{k=1}^{n_{ij}} X_{ijk} , \qquad i<j, \ i,j=1,\dots,t.$$

Furthermore, the m_{ij}'s are also independent of each other, and from the
properties of the binomial distribution we have

$$Em_{ij} = n_{ij}\pi_i/(\pi_i + \pi_j) , \qquad i<j, \ i,j=1,\dots,t, \qquad (2.3.1)$$

$$\text{var}(m_{ij}) = n_{ij}\pi_i\pi_j/(\pi_i + \pi_j)^2 \ , \quad i<j, \ i,j=1,\ldots,t, \quad (2.3.2)$$

$$\text{cov}(m_{ij},m_{kl}) = 0 \ , \quad \begin{array}{l} i<j, \ k<l, \ (i\neq k \ \text{or} \ j\neq l), \\ i,j,k,l=1,\ldots,t. \end{array} \quad (2.3.3)$$

From (2.3.1), (2.3.2), and (2.3.3), it follows that

$$E(m_{ij}m_{kl}) = (Em_{ij})(Em_{kl}) = \frac{n_{ij}\pi_i}{\pi_i + \pi_j}\frac{n_{kl}\pi_k}{\pi_k + \pi_l} \ , \quad (2.3.4)$$

and

$$Em_{ij}^2 = \text{var}(m_{ij}) + (Em_{ij})^2 = \frac{n_{ij}\pi_i\pi_j + n_{ij}^2\pi_i^2}{(\pi_i + \pi_j)^2} \ . \quad (2.3.5)$$

Under certain regularity conditions verified by Springall (1973), it follows from a theorem found in Wilks (1962, page 380) that $\sqrt{N}(\hat\beta_1-\beta_1), \ \ldots \ , \ \sqrt{N}(\hat\beta_s-\beta_s)$ is asymptotically multivariate normal, with mean vector $\underline{0}$ and variance-covariance matrix $(\lambda_{ab})^{-1}$, where

$$\lambda_{ab} = E\left[\frac{\partial \ln L}{\partial \beta_a}\frac{\partial \ln L}{\partial \beta_b}\right] \ , \quad a,b=1,\ldots,s. \quad (2.3.6)$$

We now proceed to derive the asymptotic variance-covariance matrix by deriving λ_{ab} for general a and b. From (2.2.2) we have

$$\frac{\partial \ln L}{\partial \beta_a} = \sum_{i<j}\sum \frac{x_{ai}-x_{aj}}{\pi_i + \pi_j}m_{ij}\pi_j \ + \ \sum_{j<i}\sum \frac{x_{ai}-x_{aj}}{\pi_i + \pi_j}m_{ij}\pi_j$$

$$= \sum_{i<j} \sum \left[\frac{x_{ai}-x_{aj}}{\pi_i + \pi_j} m_{ij}\pi_j + \frac{x_{aj}-x_{ai}}{\pi_i + \pi_j} (n_{ij}-m_{ij})\pi_i \right]$$

$$= \sum_{i<j} \sum \left[m_{ij}(x_{ai}-x_{aj}) - n_{ij}\pi_i \frac{x_{ai}-x_{aj}}{\pi_i + \pi_j} \right], \quad a=1,\ldots,s. \quad (2.3.7)$$

Then from (2.3.4)-(2.3.7), we find λ_{ab} to be

$$\lambda_{ab} = E \sum_{i<j} \sum \sum_{k<l} \sum \left[m_{ij}(x_{ai}-x_{aj}) - n_{ij}\pi_i \frac{x_{ai}-x_{aj}}{\pi_i + \pi_j} \right] \cdot$$

$$\left[m_{kl}(x_{bk}-x_{bl}) - n_{kl}\pi_k \frac{x_{bk}-x_{bl}}{\pi_k + \pi_l} \right]$$

$$= \left\{ \begin{array}{l} \displaystyle\sum_{\substack{i<j \\ (i \neq k \text{ or } j \neq l)}} \sum \sum_{k<l} \sum \frac{n_{ij}n_{kl}\pi_i\pi_k}{(\pi_i + \pi_j)(\pi_k + \pi_l)} (x_{ai}-x_{aj})(x_{bk}-x_{bl}) \\[3ex] + \displaystyle\sum_{i<j} \sum \frac{n_{ij}\pi_i\pi_j + n_{ij}^2\pi_i^2}{(\pi_i + \pi_j)^2} (x_{ai}-x_{aj})(x_{bk}-x_{bl}) \\[3ex] - 2\displaystyle\sum_{i<j} \sum \sum_{k<l} \sum \frac{n_{ij}\pi_i}{\pi_i + \pi_j} \frac{n_{kl}\pi_k}{\pi_k + \pi_l} (x_{ai}-x_{aj})(x_{bk}-x_{bl}) \\[3ex] + \displaystyle\sum_{i<j} \sum \sum_{k<l} \sum \frac{n_{ij}n_{kl}\pi_i\pi_k}{(\pi_i + \pi_j)(\pi_k + \pi_l)} (x_{ai}-x_{aj})(x_{bk}-x_{bl}) \end{array} \right\}$$

$$= \sum_{i<j} \sum n_{ij} \frac{\pi_i\pi_j}{(\pi_i + \pi_j)^2} (x_{ai}-x_{aj})(x_{bi}-x_{bj}), \quad a,b=1,\ldots,s. \quad (2.3.8)$$

As previously mentioned, the remaining chapters deal with finding optimal designs when the logarithm of the treatment parameters is expressed as a polynomial equation in a single variable. The various optimality criteria considered include

(i) Minimization of the variance of the maximum likelihood estimator of one parameter,

(ii) D-optimality,

(iii) Minimization of the average variance of the predicted "response", $\ln \hat{\pi}_x$,

(iv) Minimization of the maximum variance of $\ln \hat{\pi}_x$ (minimax),

(v) Minimization of the bias,

(vi) Minimization of the average mean square error of $\ln \hat{\pi}_x$ when bias is present (J-criterion),

(vii) D-optimality for a preliminary test estimator.

These are all defined in subsequent chapters.

For the case of a single quantitative independent variable, the model given in (1.5.1) becomes

$$\ln \pi_i = \sum_{k=1}^{s} \beta_k x_i^{k} \, , \qquad i=1,\ldots,t. \qquad (2.3.9)$$

Similarly, the variance-covariance matrix given in (2.3.8) becomes $(\lambda_{ab})^{-1}$, where

$$\lambda_{ab} = \sum_{i<j} \sum n_{ij} \frac{\pi_i \pi_j}{(\pi_i + \pi_j)^2} (x_i^{a} - x_j^{a})(x_i^{b} - x_j^{b}), \quad a,b=1,\ldots,s. \qquad (2.3.10)$$

Notice that the complete design problem is quite complex. It involves deciding how many levels and additionally which particular levels should be chosen. Furthermore, it must be determined how often to compare each pair of treatments.

A further complication is apparent by noticing that the variance-covariance matrix, whose elements are given in (2.3.10), depends on the true parameter values. Hence the optimal designs also depend on the parameters. The designs discussed by El-Helbawy and Bradley and Spring-all were found under the assumption that all Bradley-Terry treatment parameters are equal, or equivalently that $\beta_1 = \beta_2 = \cdots = \beta_s = 0$. In no paper of which the author is aware was the dependency of the optimal design on the true parameter values examined. As will become apparent in the following chapters, the optimal design for one set of parameter values can be quite different from the optimal design for another set of parameter values.

CHAPTER 3

MINIMIZATION OF THE VARIANCE
OF A SINGLE PARAMETER ESTIMATE

3.1 Introduction

The present chapter contains a discussion of the linear, quadratic, and cubic models. The optimality criterion considered is the minimization of $\text{var}(\hat{\beta}_i)$, where i=1,2, and 3 for the linear, quadratic, and cubic models, respectively. The linear model is discussed in Section 3.2. Designs for the quadratic model are found in Section 3.3 for $\beta_2 = 0$, β_1 variable. The general expression for $\text{var}(\hat{\beta}_2)$, given by (3.2.6), can be shown to be a continuous function of β_2, and hence the designs in Section 3.3 are locally optimal at the point $\beta_2 = 0$. Likewise, designs for the cubic model are found in Section 3.4 for $\beta_3 = 0$, β_1 and β_2 variable, and are similarly locally optimal.

3.2 Linear Model

From (2.3.9), the linear model is

$$\ln \pi_i = \beta x_i , \qquad\qquad i = 1, \ldots, t. \qquad (3.2.1)$$

Since there is only one parameter, by (2.3.10) the variance-covariance matrix becomes a simple variance, λ^{-1}, where

$$\lambda = \sum_{i<j} \sum n_{ij} \frac{e^{\beta(x_i + x_j)}}{(e^{\beta x_i} + e^{\beta x_j})^2} (x_i - x_j)^2$$

34

$$= \sum_{i<j} \sum n_{ij} \frac{(x_i - x_j)^2}{(\exp(\beta(x_i-x_j)/2) + \exp(\beta(x_j-x_i)/2))^2}$$

$$= \sum_{i<j} \sum n_{ij} \frac{d_{ij}^2}{(e^{\beta d_{ij}/2} + e^{-\beta d_{ij}/2})^2} \quad , \tag{3.2.2}$$

where $d_{ij}=x_i-x_j$. For the remainder of this dissertation, it is assumed without loss of generality that the independent variable, x, is scaled such that it is in the interval $[-1,1]$.

Although it is not necessary to treat the two cases $\beta=0$ and $\beta\neq 0$ separately, the special case $\beta=0$ is considered first because the argument used to derive the optimal designs is easier to follow for this situation. When $\beta=0$, (3.2.2) reduces to

$$\lambda = \sum_{i<j} \sum n_{ij}(x_i - x_j)^2/4 \quad . \tag{3.2.3}$$

The expression in (3.2.3) is a maximum when all the "weight" of the n_{ij} is placed onto pairs which maximize $(x_i - x_j)^2$. Clearly this implies that the design which minimizes $\text{var}(\hat{\beta})$ is a comparison of only the pair $(-1,1)$.

The designs which maximize λ when $\beta\neq 0$ are found using an argument identical to the one used for the case $\beta=0$. From (3.2.2), all of the "weight" of the n_{ij} should be placed onto pairs which maximize

$$y = d^2(e^{\beta d/2} + e^{-\beta d/2})^2 \quad , \tag{3.2.4}$$

where d is then the optimal difference between the levels in the pairs. The value of d which maximizes y in (3.2.4) is the solution to

$$\frac{\partial y}{\partial d} = \frac{2d}{(e^{\beta d/2} + e^{-\beta d/2})^2} - \frac{\beta d^2 (e^{\beta d/2} - e^{-\beta d/2})}{(e^{\beta d/2} + e^{-\beta d/2})^3} = 0 \ . \qquad (3.2.5)$$

The equality in (3.2.5) implies

$$\coth(\beta d/2) = \beta d/2 \ . \qquad (3.2.6)$$

The solution to $\coth(u) = u$ is a transcendental number, and can be shown to be approximately $u=1.1997$. Hence, from (3.2.6) and the fact that $|d|$ must be less than or equal to 2, the design which minimizes $\mathrm{var}(\hat{\beta})$ for a given value of β is

$$|d| = |x_i - x_j| = \min(2, 2.3994/|\beta|) \ . \qquad (3.2.7)$$

It is now shown that the value of $|d|$ given in (3.2.7) is in fact a maximum for y, and consequently for λ. By (3.2.5), the second partial derivative of y with respect to d is given by

$$\frac{\partial^2 y}{\partial d^2} (e^{\beta d/2} + e^{-\beta d/2})^2 =$$

$$2 - 4d\beta \cdot \tanh(\beta d/2) - \beta^2 d^2/2 + \frac{3}{2}\beta^2 d^2 \tanh^2(\beta d/2) \ . \qquad (3.2.8)$$

The second partial derivative $\partial^2 y/\partial d^2$ is negative if and only if the expression in (3.2.8) is negative. From (3.2.6), the equality

$$2 = \beta d \cdot \tanh(\beta d/2) \qquad (3.2.9)$$

holds at $d=2.3994/|\beta|$, which is then substituted into (3.2.8) to give

$$\left. \frac{\partial^2 y}{\partial d^2} \right|_{\text{opt. } d} = 2 - 8 - \beta^2 d^2/2 + 6 = -\beta^2 d^2/2 < 0 \ . \qquad (3.2.10)$$

Because there is only one solution to (3.2.6) where d>0, the result in

(3.2.10) implies that the solution of $|d|$ given in (3.2.7) does in fact

minimize the variance of $\hat{\beta}$.

By (3.2.7) the optimal design is a comparison of only the pair

(-1,1) when β <1.2. It can be shown that $P(1 \to -1)=.92$ when $\beta=1.2$.

Because of this relatively large preference probability of choosing one

endpoint over the other, in most practical situations β will be less

than 1.2. So in these cases the optimal design is a comparison of only

the endpoints of the experimental region.

The optimal designs given by (3.2.7) are also optimal for other

criteria. In particular, they are optimal for criteria (ii)-(iv) intro-

duced in Section 2.3. These criteria are formally defined in Chapter 4,

but a brief explanation for their equivalence in the case of a linear

model follows. The equivalence of the criterion being discussed in the

present section to D-optimality follows from the fact that the deter-

minant of a 1×1 matrix is simply the only element in the matrix. The

equivalence to the other two criteria follows from the fact that

$$\text{var}(\ln \hat{\pi}_x) \;=\; x^2 \text{var}(\hat{\beta}) \quad . \tag{3.2.11}$$

From (3.2.11), criterion (iii) is, in this case, the minimization of

$$\int_{-1}^{1} x^2 \text{var}(\hat{\beta}) \; dx \;=\; \frac{2}{3} \text{var}(\hat{\beta}) \quad ,$$

and criterion (iv) is the minimization of

$$\max_{x \varepsilon [-1,1]} x^2 \text{var}(\hat{\beta}) \;=\; \text{var}(\hat{\beta}) \quad .$$

3.3 Quadratic Model

The quadratic model is

$$\ln \pi_i = \beta_1 x_i + \beta_2 x_i^2 , \qquad\qquad i=1,\ldots,t. \qquad\qquad (3.3.1)$$

From (2.3.10), the asymptotic variance-covariance matrix of the maximum likelihood estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ is $(\lambda_{ab})^{-1}$, where

$$\lambda_{11} = \sum_{i<j}\sum n_{ij}\phi_{ij}(x_i - x_j)^2 , \qquad\qquad (3.3.2)$$

$$\lambda_{12} = \sum_{i<j}\sum n_{ij}\phi_{ij}(x_i - x_j)(x_i^2 - x_j^2) , \qquad\qquad (3.3.3)$$

$$\lambda_{22} = \sum_{i<j}\sum n_{ij}\phi_{ij}(x_i^2 - x_j^2)^2 , \qquad\qquad (3.3.4)$$

and where

$$\phi_{ij} = \phi(x_i,x_j) = \frac{\exp(\beta_1(x_i + x_j) + \beta_2(x_i^2 + x_j^2))}{(\exp(\beta_1 x_i + \beta_2 x_i^2) + \exp(\beta_1 x_j + \beta_2 x_j^2))^2} ,$$

$$i<j, \ i,j=1,\ldots,t. \qquad\qquad (3.3.5)$$

Minimization of $\mathrm{var}(\hat{\beta}_2)$
_____

Consider the test of the hypothesis

$$H_o: \ \beta_2 = 0 , \ \ \beta_1 \text{ unspecified},$$

versus

$$H_a: \ \beta_2 \neq 0 , \ \ \beta_1 \text{ unspecified}.$$

Minimizing the asymptotic variance of $\hat{\beta}_2$ would equivalently be maximizing the asymptotic power of the test. The variance of $\hat{\beta}_2$ is the lower right-hand element of the 2×2 matrix $(\lambda_{ab})^{-1}$. That is,

$$\text{var}(\hat{\beta}_2) = \frac{\lambda_{11}}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \; . \tag{3.3.6}$$

Under the null hypothesis, (3.3.5) becomes

$$\phi_{ij} = \frac{e^{\beta_1(x_i + x_j)}}{(e^{\beta_1 x_i} + e^{\beta_1 x_j})^2} \; , \qquad i<j, \; i,j=1,\ldots,t. \tag{3.3.7}$$

As was pointed out in Section 3.1, in the present section (3.3.6) is minimized for $\beta_2=0$. This produces locally optimal designs. Assuming $\beta_2=0$ greatly simplifies the design problem as will be seen in the remainder of the section.

The end result of Lemmas 3.1-3.2 and Theorems 3.1-3.3 is a set of optimal designs which minimize $\text{var}(\hat{\beta}_2)$ under the null hypothesis for various choices of the linear parameter. Hereafter, "optimal design" will refer to the design which is optimal for the criterion being discussed in the section.

The following two definitions are useful in the discussion that follows.

Definition 3.1. The pair of treatments $(-x_1,-x_2)$ is defined to be the symmetric counterpart of the pair (x_1,x_2).

Definition 3.2. A design is said to be symmetric if each pair in the design is compared as often as its respective symmetric counterpart.

<u>Lemma 3.1</u>. Let $\phi(x_i, x_j)$ denote ϕ_{ij} as given by (3.3.7). Then $\phi(x_i, x_j) = \phi(-x_i, -x_j)$, for any x_i and x_j.

Proof.

$$\phi(x_i, x_j) = \frac{e^{\beta_1(x_i + x_j)}}{(e^{\beta_1 x_i} + e^{\beta_1 x_j})^2} \cdot \frac{e^{-2\beta_1(x_i + x_j)}}{e^{-2\beta_1(x_i + x_j)}}$$

$$= \frac{e^{\beta_1(-x_i - x_j)}}{(e^{\beta_1(-x_j)} + e^{\beta_1(-x_i)})^2}$$

$$= \phi(-x_i, -x_j) . \qquad †$$

<u>Theorem 3.1</u>. The optimal design is a symmetric design.

Proof. Notice that both of the following hold:

(i) $(x_j - x_i)^2 = ((-x_j) - (-x_i))^2$,

(ii) $(x_j^2 - x_i^2) = ((-x_j)^2 - (-x_i)^2)$.

From (3.3.2), (3.3.4), Lemma 3.1, and the above, it is clear that λ_{11} and λ_{22} are invariant under changes of any number of comparisons of pairs to comparisons of their symmetric counterparts.

Let D_1 denote any particular design, and let $\text{var}_{D_1}(\hat{\beta}_2)$ be the variance of $\hat{\beta}_2$ for Design D_1. Also, let the ith pair in D_1 be denoted (x_i, x_i'). The sets of comparisons $\{(x_i, x_i'), (-x_i, -x_i')\}$ are considered in sequence, and without loss of generality, it is assumed that

$$n_{x_i, x_i'} - n_{-x_i, -x_i'} = d_i \geq 0 ,$$

for all i. Notice that it could be true that $n_{-x_i, -x_i'} = 0$.

Design D_2 is formed from Design D_1 by changing $d_i/2$ of the compar-
isons of the pair (x_i, x_i') to comparisons of the pair $(-x_i, -x_i')$. This
results in $(n_{x_i, x_i'} + n_{-x_i, -x_i'})/2$ comparisons of each of the two pairs
in the ith set. This change is made for each set of comparisons, there-
by defining Design D_2 to be a symmetric design. The total number of
comparisons, N, is left unchanged.

By the first paragraph of this proof, the values of λ_{11} and λ_{22}
are the same for the two designs. By (3.3.3), $\lambda_{12}=0$ for Design D_2,
and hence λ_{12}^2 is a minimum. Therefore, by (3.3.6),

$$\text{var}_{D_2}(\hat{\beta}_2) \leq \text{var}_{D_1}(\hat{\beta}_2) \quad . \tag{3.3.8}$$

Note that the variances are equal if $\lambda_{12}=0$ for Design D_1, and it
is possible that $\lambda_{12}=0$ for a nonsymmetric design. However, (3.3.8)
shows that no design can "do better" than a symmetric design. Also
note that in practice, d_i must be an even integer in order for the
n_{ij}'s to be integers.

So the search for optimal designs in the present section will hence-
forth be restricted to the class of all symmetric designs.

Because $\lambda_{12}=0$ for a symmetric design, (3.3.6) reduces to

$$\text{var}(\hat{\beta}_2) = 1/\lambda_{22} \quad . \tag{3.3.9}$$

So it is now necessary to find designs which maximize λ_{22}. Notice that
λ_{22} depends on β_1. As previously mentioned, the optimal design also
depends on the value of β_1.

Theorem 3.2. The optimal design is a design in which the only compar-
isons made are comparisons of the pairs (x,1) and (-1,-x) an equal num-
ber of times, for some x. Furthermore, $x_. \geq 0$.

Proof. First of all, the notation used for λ_{22}, which is defined by
(3.3.4) and (3.3.7), is changed. The N total comparisons are arbitrar-
ily ordered, and the ith pair is then defined to be (x_i, x_i'). The vari-
ables y_i and z_i are defined to be

$$y_i = x_i - x_i' , \qquad z_i = x_i + x_i' , \qquad i=1,\ldots,N.$$

Then from (3.3.7),

$$\phi_i = \frac{e^{\beta_1(x_i + x_i')}}{(e^{\beta_1 x_i} + e^{\beta_1 x_i'})^2} \cdot \frac{e^{-\beta_1(x_i + x_i')}}{e^{-\beta_1(x_i + x_i')}}$$

$$= \frac{1}{(e^{\beta_1 y_i/2} + e^{-\beta_1 y_i/2})^2} , \qquad i=1,\ldots,N.$$

Notice that ϕ_i is a function of y_i but not of z_i. From (3.3.4), λ_{22} is
rewritten as

$$\lambda_{22} = \sum_{i=1}^{N} \phi_i y_i^2 z_i^2 .$$

Let D be any design. By holding y_i fixed, we examine what changes
in z_i^2 will increase λ_{22}. Clearly if $|z_i|$ is as large as possible for
every i, then λ_{22} is also as large as possible for the particular set
of y_i's found in D. Without loss of generality, it is assumed that
$x_i < x_i'$. Then if either $x_i = -1$ or $x_i' = 1$, $|z_i|$ is maximized. Therefore,

all comparisons must involve either x=-1 or x=1. Additionally, using
an argument similar to one used in Section 3.2, it is clear from the
expression for λ_{22} given in (3.3.4) that x=1 should be compared with
one and only one level of x, namely the value of x which maximizes
$\phi(1,x) \cdot (1-x^2)^2$. This result in conjunction with Theorem 3.1 proves the
first part of the theorem.

We can now rewrite λ_{22} simply as

$$\lambda_{22} = N\phi(1,x) \cdot (1-x^2)^2 \ .$$

It remains to be proven that the optimal value of x is such that $x \geq 0$.

The variable $\psi(x)$ is defined to be the positive square root of
λ_{22}/N, i.e.

$$\psi(x) = \frac{e^{\beta_1(1 + x)/2}}{e^{\beta_1} + e^{\beta_1 x}} (1 - x^2) \ . \tag{3.3.10}$$

Since (3.3.10) and the fact that $x \in [-1,1]$ imply $\psi(x) \geq 0$, maximizing
λ_{22} is equivalent to maximizing $\psi(x)$.

It will now be shown that $\psi(x) \geq \psi(-x)$, $x \geq 0$. The following five
inequalities are equivalent:

$$\psi(x) \geq \psi(-x) \tag{3.3.11}$$

$$\frac{e^{\beta_1(1 + x)/2}}{e^{\beta_1} + e^{\beta_1 x}} \geq \frac{e^{\beta_1(1 - x)/2}}{e^{\beta_1} + e^{-\beta_1 x}} \tag{3.3.12}$$

$$e^{\frac{1}{2}\beta_1(1+x)+\beta_1} + e^{\frac{1}{2}\beta_1(1+x)-\beta_1 x} \geq e^{\frac{1}{2}\beta_1(1-x)+\beta_1} + e^{\frac{1}{2}\beta_1(1-x)+\beta_1 x} \tag{3.3.13}$$

$$e^{\beta_1(1 + x)/2} + e^{-\beta_1(1 + x)/2} \geq e^{\beta_1(1 - x)/2} + e^{-\beta_1(1 - x)/2} \tag{3.3.14}$$

$$\cosh(u + \beta_1 x) \geq \cosh(u) , \tag{3.3.15}$$

where

$$\cosh(u) = (e^u + e^{-u})/2 ,$$

$$u = \beta_1(1 - x)/2 .$$

Now $\cosh(u)$ is a monotonically increasing function of u for $u \geq 0$. Furthermore, it can be assumed that $\beta_1 \geq 0$ because Theorem 3.3 proves that the optimal designs for $\beta_1 = \pm\beta_{10}$ are the same. Or, equivalently, the remainder of this proof can be repeated for the case $\beta_1 < 0$. Then $x \geq 0$ and $\beta_1 \geq 0$ implies $\beta_1 x \geq 0$, and hence by the equivalence of (3.3.11) and (3.3.15), $\psi(x) \geq \psi(-x)$ for all $x \geq 0$. This completes the proof of the second part. †

The next theorem shows that optimal designs only need to be found for nonnegative values of β_1.

<u>Theorem 3.3</u>. The optimal design for $\beta_1 = -\beta_{10}$ is the same as the optimal design for $\beta_1 = \beta_{10}$.

Proof. Recall that the optimal design is found by finding the value of x which maximizes $\psi(x)$ from the proof of Theorem 3.2. Since $\psi(x)$ is actually a function of β_1 and x, we write

$$\psi(\beta_1, x) = \frac{e^{\beta_1(1 + x)/2}}{e^{\beta_1} + e^{\beta_1 x}} (1 - x^2) .$$

The theorem is proved by observing that

$$\psi(\beta_1,x) = \frac{e^{\beta_1(1 + x)/2}}{e^{\beta_1} + e^{\beta_1 x}} (1 - x^2) \cdot \frac{e^{-\beta_1(1 + x)}}{e^{-\beta_1(1 + x)}}$$

$$= \frac{e^{-\beta_1(1 + x)/2}}{e^{-\beta_1 x} + e^{-\beta_1}} (1 - x^2)$$

$$= \psi(-\beta_1,x) . \qquad \dagger$$

All that remains to be done is to find the value of x that maximizes $\psi(x)$ for various choices of $\beta_1 \geq 0$. Unfortunately, a closed form solution of x as a function of β_1 does not exist. However when $\beta_1=0$, the optimal value of x can be analytically derived, and hence this situation is considered first.

Suppose $\beta_1=0$. Then from (3.3.10), $\psi(x) = (1 - x^2)/2$. Clearly this is maximized by x=0, and hence from Theorems 3.1 and 3.2, the optimal design is an equal number of comparisons of only the pairs (0,1) and (-1,0). For the case $\beta_1 \neq 0$, the grid method described in the next paragraph was used to obtain the results presented in Table 3.1 and Figure 3.1.

The grid method utilized first calculated $\psi(x)$ for x=0(.1)1, and the best value of x, say x_1, out of these 11 choices was found. Then $\psi(x)$ was calculated for the 11 values x=$(x_1-.1),(x_1-.08), \ldots ,$ $(x_1+.1)$, and again the best value of x was found. This procedure was repeated until the optimal value of x was found accurate to the fourth

decimal place. The entire procedure was done separately for each value of $|\beta_1|$.

The following lemma proves that the values of x which are found via the grid search are absolute maximums for $\psi(x)$.

Lemma 3.2. Exactly one local maximum (and hence it is the absolute maximum) exists for $\psi(x)$ in the interval $[0,1]$.

Proof. By Theorem 3.3, only the case $\beta_1 > 0$ needs to be considered. From (3.3.10), the first partial derivative of $\psi(x)$ with respect to x is

$$\psi'(x) = \frac{e^{\beta_1(1 + x)/2}}{(e^{\beta_1} + e^{\beta_1 x})^2} \left[e^{\beta_1}(\beta_1(1 - x^2)/2 - 2x) - e^{\beta_1 x}(\beta_1(1 - x^2)/2 + 2x) \right]. \quad (3.3.16)$$

It can be seen from (3.3.16) that $\psi'(0) > 0$ and $\psi'(1) < 0$. Hence $\psi'(x) = 0$ for at least one $x \epsilon [0,1]$, and it remains to be shown that $\psi'(x) = 0$ for exactly one $x \epsilon [0,1]$. (An attempt was made to show $\psi''(x)$ is negative for all β_1 and $x \epsilon [0,1]$, but it turned out not to be true. In particular, a graph of $\psi(x)$ as a function of x for $\beta_1 = 5$ indicated $\psi''(x) > 0$ for values of x near 0.)

From (3.3.16), $\psi'(x) = 0$ if and only if

$$(\beta_1(1 - x^2)/2 - 2x) = e^{-\beta_1(1 - x)}(\beta_1(1 - x^2)/2 + 2x). \quad (3.3.17)$$

Notice that the functions $f_1(x) = (\beta_1(1 - x^2)/2 - 2x)$ and $f_2(x) = (\beta_1(1 - x^2)/2 + 2x)$ are reflections of each other about the axis x=0, and they are quadratic in x. Hence they intersect exactly

once at x=0. Also, for fixed β_1, f_1 and f_2 have maximums at $-2/\beta_1$ and $2/\beta_1$, respectively. Let $g_1(x)$ be the right-hand side of (3.3.17). It was shown in the preceding paragraph that $f_1(x)$ and $g_1(x)$ intersect at least once in the interval $[0,1]$. To determine that they intersect exactly once, it is sufficient to show that $\partial f_1(x)/\partial x < \partial g_1(x)/\partial x$.

From (3.3.17),

$$\frac{\partial f_1(x)}{\partial x} = -\beta_1 x - 2 ,$$

$$\frac{\partial g_1(x)}{\partial x} = \beta_1 e^{-\beta_1(1-x)}(\beta_1(1-x^2)/2 + 2x) + e^{-\beta_1(1-x)}(-\beta_1 x + 2)$$

$$= e^{-\beta_1(1-x)}(\beta_1^2(1-x^2)/2 + \beta_1 x + 2) .$$

For $\beta_1 > 0$ and $x \in [0,1]$, it is clear that $\partial f_1(x)/\partial x < 0$ and $\partial g_1(x)/\partial x > 0$, which completes the proof. †

Table 3.1 presents the optimal value of x found from the grid search described immediately prior to Lemma 3.2, for various choices of $|\beta_1|$. Figure 3.1 is a graph of the optimal value of x versus $|\beta_1|$. The table includes relatively large values of $|\beta_1|$ to illustrate that $x \uparrow 1$ as $|\beta_1| \uparrow \infty$. Recall that for a given value of $|\beta_1|$, the optimal design is a comparison of $(-1,-x)$ and $(x,1)$ an equal number of times, where x is found in Table 3.1 or Figure 3.1.

Table 3.1 indicates that the optimal design when $\beta_1 = 0$ involves only 3 levels, x=-1,0,1. However, when $\beta_1 \neq 0$, the optimal design requires 4 levels, $x = -1, -x_o, x_o, 1$, for some x_o. In certain experimental situations it may be significantly more laborious to make four "batches"

Table 3.1.   Designs which minimize var($\hat{\beta}_2$)

$|\beta_1|$	Optimal x	Relative efficiency
0	0	1
.05	.0003	1
.10	.0012	1
.15	.0028	1
.20	.0050	1
.25	.0077	.9999
.30	.0110	.9998
.35	.0150	.9995
.40	.0194	.9992
.45	.0243	.9988
.50	.0297	.9982
.55	.0356	.9974
.60	.0419	.9963
.65	.0486	.9950
.70	.0556	.9935
.75	.0630	.9915
.80	.0708	.9893
.85	.0787	.9866
.90	.0869	.9836
.95	.0953	.9801
1.0	.1039	.9761
1.1	.1216	.9668
1.2	.1396	.9556
1.3	.1580	.9426
1.4	.1764	.9269
1.5	.1948	.9097
1.6	.2130	.8905
1.7	.2310	.8696
1.8	.2487	.8472
1.9	.2660	.8233
2.0	.2829	.7983
3.0	.4269	.5242
4.0	.5295	.2974
5.0	.6031	.1546
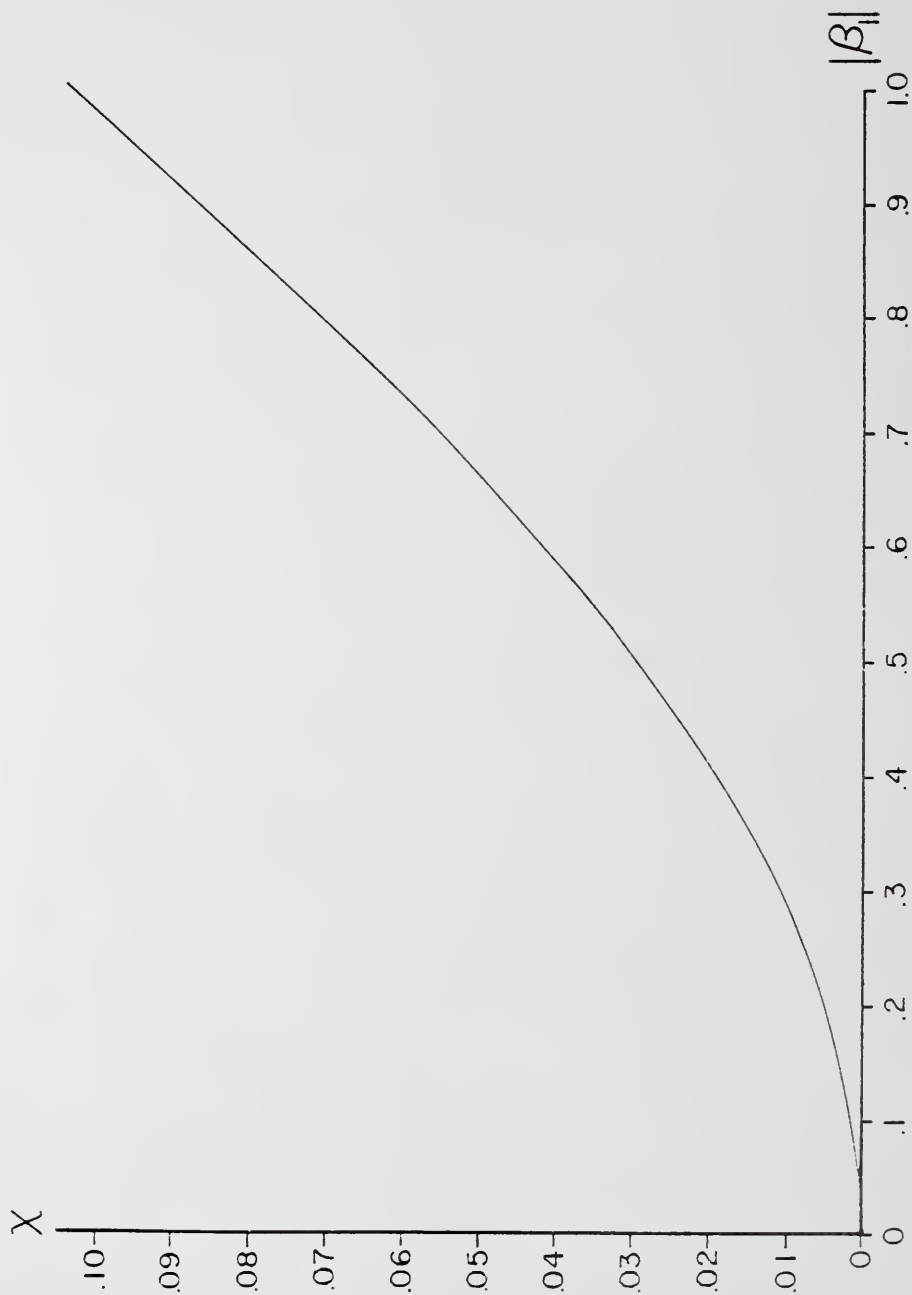
Figure 3.1. Designs which minimize var($\hat{\beta}_2$)

than it is to make just three. For this reason, and because a priori

information about β_1 may not be available, asymptotic relative efficien-

cies were calculated for the design $\{n_{-1,0} = n_{0,1} = N/2\}$ to the optimal

designs found in the table. The relative efficiency of the first design

to the second (optimal) design is defined to be

$$\text{R.E.} \quad = \quad \frac{\text{var}_2(\hat{\beta}_2)}{\text{var}_1(\hat{\beta}_2)}$$

$$= \quad \frac{e^{\beta_1}/(e^{\beta_1} + 1)^2}{e^{\beta_1(1 + x)}(1 - x^2)^2/(e^{\beta_1} + e^{\beta_1 x})^2} \quad .$$

The relative efficiencies were enumerated for various choices of β_1,

and are presented next to the optimal designs in Table 3.1.

Equivalence of El-Helbawy and Bradley's method and minimizing the variance

An application of El-Helbawy and Bradley's method for finding opti-

mal designs was briefly discussed in Section 1.4. It will now be prov-

en that their optimality criterion is equivalent to minimizing $\text{var}(\hat{\beta}_k)$,

any k. First of all, notation and results of El-Helbawy and Bradley's

paper which are necessary for the proof of Theorem 3.5 are presented.

Recall that El-Helbawy and Bradley considered linear contrasts of

$\gamma_i = \ln \pi_i$, $i=1, \ldots ,t$. The matrices $\underline{B}_m$ and $\underline{B}_n$ are defined as m×t and

n×t, respectively, with zero-sum, orthonormal rows, such that

$0 \leq m < m+n \leq t-1$. The null hypothesis is

$$H_0: \quad \underline{B}_n \underline{\gamma}(\underline{\pi}) = \underline{0}_n \quad ,$$

and the alternative hypothesis is

$$H_a: \quad \underline{B}_n \underline{\gamma}(\underline{\pi}) \neq \underline{0}_n \quad ,$$

where $\underline{\pi}$ and $\underline{\gamma}(\underline{\pi})$ are column vectors with i^{th} elements π_i and γ_i, respectively. The rows of $\underline{B}_m$ represent the linear contrasts assumed to be null. Note that $\underline{B}_m$ need not exist (i.e. m=0).

The vector $\underline{\pi}_o$ is a value of $\underline{\pi}$ which satisfies the null hypothesis and the constraints, where $\underline{\pi}$ is the true parameter vector.

The matrix $\underline{B}$ is defined to be the t×t orthonormal matrix

$$\underline{B} \;=\; \begin{bmatrix} \underline{1}'/\sqrt{t} \\[2mm] \underline{B}_m \\[2mm] \underline{B}_m^* \end{bmatrix} \;, \tag{3.3.18}$$

where $\underline{B}_m^*$ is any (t-m-1)×t matrix such that $\underline{BB}' = \underline{I}_t$, and $\underline{1}'=(1, \ldots ,1)$, the t-dimensional unit vector.

The following theorem is extracted from El-Helbawy and Bradley (1978, Theorem 2, page 833).

Theorem 3.4. Let $\underline{p}$ be the maximum likelihood estimator of $\underline{\pi}$. Under the assumption of connectedness (Assumption 1.2), and by the assumption of strictly positive elements of $\underline{\pi}$, which has already been inherently assumed since ln(0) is undefined, $\sqrt{N}(\underline{\gamma}(\underline{p}) - \underline{\gamma}(\underline{\pi}))$ has a limiting distribution function that is singular, t-variate normal in a space of (t-m-1) dimensions, with zero mean vector and dispersion matrix

$$\Sigma(\underline{\pi}) \;=\; \underline{B}_m^{*\,\prime} (\underline{C}(\underline{\pi}))^{-1} \underline{B}_m^* \;, \tag{3.3.19}$$

where

$$\underline{C}(\underline{\pi}) = \underline{B}^*_m \underline{\Lambda}(\underline{\pi}) \underline{B}^*_m{}' \, , \qquad (3.3.20)$$

and $\underline{\Lambda}(\underline{\pi})$ is the t×t matrix with elements

$$\Lambda_{ii}(\underline{\pi}) = \pi_i \sum_{\substack{j=1 \\ j \neq i}}^{t} \mu_{ij} \pi_j / (\pi_i + \pi_j)^2 \, , \qquad i=1,\dots,t,$$

$$\qquad (3.3.21)$$

$$\Lambda_{ij}(\underline{\pi}) = -\mu_{ij} \pi_i \pi_j / (\pi_i + \pi_j)^2 \, , \quad i \neq j, \ i,j=1,\dots,t,$$

and where μ_{ij} is defined in Assumption 1.2.

El-Helbawy and Bradley's optimality criterion is the maximization of the noncentrality parameter, λ^2, for the asymptotic chi-square distribution of the likelihood ratio statistic. The noncentrality parameter is

$$\lambda^2 = \underline{\delta}' \Sigma_o^{-1} \underline{\delta} \, , \qquad (3.3.22)$$

where $\underline{\delta}$ is a vector of constants defined in their paper, Σ_o is the leading principal matrix of order n of $\underline{C}_o^{-1}$, $\underline{B}_n$ provides the first n rows of $\underline{B}^*_m$, and

$$\underline{C}_o = \underline{B}^*_m \underline{\Lambda}(\underline{\pi}_o) \underline{B}^*_m{}' \, . \qquad (3.3.23)$$

A theorem proving the equivalence of the two optimality criteria is now presented.

Theorem 3.5. Consider the model $\ln \pi_x = \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k$. Suppose we are interested in testing, for some fixed i,

$$H_o : \ \beta_i = 0 \qquad \qquad \text{versus} \qquad \qquad H_a : \ \beta_i \neq 0 \, .$$

Then the design which minimizes $\text{var}(\hat{\beta}_i)$ is identical to the design which maximizes the noncentrality parameter from the likelihood ratio test. (Note that it does not matter whether or not it is assumed that some of the other β_j's are 0.)

Proof. Orthogonal polynomials can be derived for any spacing of the independent variable. Let these contrasts be the rows of $\underline{B}$. The 1×t matrix $\underline{B}_{n=1}$ is the orthogonal polynomial corresponding to the parameter β_i being tested.

From the method used to find orthogonal polynomials, it follows that

$$\beta_i = c\underline{B}_{n=1}\underline{\gamma}(\underline{\pi}) ,$$

where c is a constant. So minimizing $\text{var}(\hat{\beta}_i)$ is equivalent to minimizing $\text{var}(\underline{B}_{n=1}\underline{\gamma}(\underline{p}))$. But from Theorem 3.4,

$$\text{var}(\underline{B}_{n=1}\underline{\gamma}(\underline{p})) = \underline{B}_{n=1}(\underline{B}_m^{*\prime}(\underline{C}(\underline{\pi}))^{-1}\underline{B}_m^{*})\underline{B}_{n=1}'$$

$$= (1,0, \ldots ,0) \; (\underline{C}(\underline{\pi}))^{-1} \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

$$= ((\underline{C}(\underline{\pi}))^{-1})_{11} .$$

From the definition of λ^2, Σ_o is the leading principal matrix of order 1 of $\underline{C}_o^{-1}$, i.e.

$$\Sigma_o = ((\underline{C}(\underline{\pi}))^{-1})_{11} .$$

In this case, $\underline{\delta}$ is a scalar constant, and so maximizing λ^2 is equivalent to minimizing $\Sigma_o = \text{var}(\hat{\beta}_i)$. †

Example 3.1. This example illustrates the equivalence of the two methods presented in Theorem 3.5. It is assumed that the true model is a quadratic model, $\ln \pi_x = \beta_1 x + \beta_2 x^2$. Also, the design is restricted to have the three design points x=-1,0,1.

Table 3.1 indicates that the design which minimizes the variance of $\hat{\beta}_2$ is $\{n_{-1,0} = n_{0,1} = N/2\}$.

The design problem is now approached using the method presented by El-Helbawy and Bradley. The null hypothesis is

$$H_o: \quad \beta_2 = 0 \ , \quad \beta_1 \text{ unknown.}$$

Using orthogonal polynomials, this hypothesis is equivalent to

$$H_o: \quad \frac{(-1 \quad 2 \quad -1)}{\sqrt{6}} \underline{\gamma}(\underline{\pi}) = 0 \ ,$$

where $\underline{\gamma}'(\underline{\pi}) = (\ln \pi_{-1}, \ln \pi_0, \ln \pi_1)$.

Note that El-Helbawy and Bradley's constraint, that the $\ln \pi_i$'s sum to zero, poses no problem. Instead of constraining π_0 to be equal to 1, an intercept parameter, β_0, could be introduced to the model. The model would then become

$$\ln \pi_x = \beta_0 + \beta_1 x + \beta_2 x^2 \ ,$$

with the constraint

$$\sum_x \ln \pi_x = 0 \ .$$

It can be shown in this particular case that $\beta_0 = -\frac{2}{3}\beta_2$.

For a definition of the following matrices and notation, see (3.3.18)-(3.3.23). For the example under discussion,

$$\underline{B}_{n=1} = (-1 \quad 2 \quad -1)/\sqrt{6} \, ,$$

$$\underline{B}_m^* = \begin{bmatrix} (-1 & 2 & -1)/\sqrt{6} \\ (-1 & 0 & 1)/\sqrt{2} \end{bmatrix} \, ,$$

$$\underline{\Lambda}(\underline{\pi}_o) = \frac{1}{4} \begin{bmatrix} \mu_{12} + \mu_{13} & -\mu_{12} & -\mu_{13} \\ -\mu_{12} & \mu_{12} + \mu_{23} & -\mu_{23} \\ -\mu_{13} & -\mu_{23} & \mu_{13} + \mu_{23} \end{bmatrix} \, ,$$

where $\underline{\pi}_o{}' = (1,1,1)$, i.e. $\beta_1 = \beta_2 = 0$. Note that $\underline{\pi}_o$ can be any vector which satisfies the null hypothesis and the constraint.

The matrix $\underline{C}_o$ is

$$\underline{C}_o = \underline{B}_m^* \underline{\Lambda}(\underline{\pi}_o) \underline{B}_m^{*\prime}$$

$$= \frac{1}{4} \begin{bmatrix} \frac{3}{2}(\mu_{12} + \mu_{23}) & \frac{3}{\sqrt{12}}(\mu_{12} - \mu_{23}) \\ \frac{3}{\sqrt{12}}(\mu_{12} - \mu_{23}) & \frac{1}{2}(\mu_{12} + 4\mu_{13} + \mu_{23}) \end{bmatrix} \, ,$$

implying

$$\underline{C}_o^{-1} = \begin{bmatrix} \frac{1}{2}(\mu_{12} + 4\mu_{13} + \mu_{23}) & -\frac{3}{\sqrt{12}}(\mu_{12} - \mu_{23}) \\ -\frac{3}{\sqrt{12}}(\mu_{12} - \mu_{23}) & \frac{3}{2}(\mu_{12} + \mu_{23}) \end{bmatrix} / |\underline{C}_o| \, .$$

Then since Σ_o is the upper left-hand element of $\underline{C}_o^{-1}$,

$$\Sigma_o = \frac{\mu_{12} + 4\mu_{13} + \mu_{23}}{2|\underline{C}_o|} \, ,$$

implying

$$\Sigma_o^{-1} = \frac{2|\underline{c}_o|}{\mu_{12} + 4\mu_{13} + \mu_{23}}$$

$$= \frac{3}{8} \frac{\mu_{12}\mu_{13} + \mu_{12}\mu_{23} + \mu_{13}\mu_{23}}{\mu_{12} + 4\mu_{13} + \mu_{23}} . \tag{3.3.24}$$

The design which maximizes the chi-square noncentrality parameter from the likelihood ratio test is the design which maximizes Σ_o^{-1}.

By the definition of the μ_{ij}, $\mu_{23} = 1 - \mu_{12} - \mu_{13}$. So Σ_o^{-1} can be rewritten as

$$\Sigma_o^{-1} = \frac{\mu_{12}\mu_{13} + (\mu_{12} + \mu_{13})(1 - \mu_{12} - \mu_{13})}{\mu_{12} + 4\mu_{13} + (1 - \mu_{12} - \mu_{13})}$$

$$= \frac{\mu_{12}(1 - \mu_{12} - \mu_{13}) + (\mu_{13} - \mu_{13}^2)}{3\mu_{13} + 1} . \tag{3.3.25}$$

Suppose that μ_{13} is temporarily held constant. It is then claimed that choosing $\mu_{12} = (1 - \mu_{13})/2$ maximizes Σ_o^{-1} for the particular μ_{13}. From (3.3.25), the expression $\mu_{12}((1 - \mu_{13}) - \mu_{12})$ must be maximized. Because of the constraint $0 \le \mu_{12} \le 1-\mu_{13}$, it is clear that $\mu_{12}((1 - \mu_{13}) - \mu_{12})$ is maximized when $\mu_{12} = (1 - \mu_{13})/2$, which proves the claim. The constraint among the μ_{ij}'s in addition to the above claim implies that the optimal design must be a design such that

$$\mu_{12} = \frac{1 - \mu_{13}}{2} = \mu_{23} . \tag{3.3.26}$$

The scalar Σ_o^{-1} is now rewritten as a function of μ_{12} alone. From (3.3.24) and (3.3.26),

$$\Sigma_o^{-1} = \frac{3}{8} \frac{2\mu_{12}(1 - 2\mu_{12}) + \mu_{12}^2}{2\mu_{12} + 4(1 - 2\mu_{12})}$$

$$= \frac{3}{8} \frac{2\mu_{12} - 3\mu_{12}^2}{4 - 6\mu_{12}}$$

$$= \frac{3\mu_{12}}{16} .$$

Clearly Σ_o^{-1} is maximized by choosing μ_{12} as large as possible. By (3.3.26), this would make $\mu_{12} = \mu_{23} = .5$, and $\mu_{13} = 0$. So this optimal design is in fact the same as the design which minimizes $\text{var}(\hat{\beta}_2)$.   †

An attempt was made to find the optimal design for the test of hypothesis

$$H_o: \quad \beta_1 = \beta_2 = 0 \qquad \text{versus} \qquad H_a: \quad \text{not } H_o$$

using El-Helbawy and Bradley's methodology. When simultaneously testing more than one parameter, or equivalently more than one treatment contrast, a constant value, δ_i, must be chosen for each parameter. The theory behind the optimal designs includes an infinite sequence of values of $\underline{\pi}$ satisfying H_a, which are converging to a vector $\underline{\pi}_o$ satisfying H_o. When the treatment contrasts are for factorial treatments, which is the case they discussed extensively, the relative sizes of the δ_i are said to be equivalent to the relative importance placed on each contrast being tested in H_o. Hence, El-Helbawy and Bradley usually chose $\delta_1 = \delta_2 = \cdots = \delta_n$.

In the present situation, choosing δ_1 and δ_2 restricts the alternative parameter values in the infinite sequence mentioned in the preceding paragraph to be a proper subset of the alternative parameters satisfying H_a. For example, $\delta_1 = \delta_2$ implies that $\beta_1 = \beta_2$, and hence the alternative models are $\ln \pi_x = \beta x + \beta x^2$, with $\beta \to 0$ as $N \to \infty$. Choosing different δ_i can result in a different design because the maximization of the noncentrality parameter depends on the relative sizes of the δ_i. Additionally, the optimal designs actually found after choosing the relative sizes of δ_1 and δ_2 were found to be unreasonable (e.g. for $\delta_1 = \delta_2$, the optimal design is a comparison of only the pair $(-1,0)$). Chapter 4 presents other optimality criteria for fitting a quadratic model.

3.4   Cubic Model

The cubic model is

$$\ln \pi_i = \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3, \qquad i=1,\ldots,t. \qquad (3.4.1)$$

The asymptotic variance-covariance matrix of the maximum likelihood estimators, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ is $(\lambda_{ab})^{-1}$, where λ_{ab} is given by (2.3.10). The inverse of the variance-covariance matrix is partitioned as

$$(\lambda_{ab}) = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{12} & \lambda_{22} & \lambda_{23} \\ \hline \lambda_{13} & \lambda_{23} & \lambda_{33} \end{bmatrix}.$$

Then by a result found in Graybill (1969, page 165), the variance of

$\hat{\beta}_3$ is

$$\text{var}(\hat{\beta}_3) = (\lambda_{ab})_{33}^{-1}$$

$$= \left[\lambda_{33} - (\lambda_{13} \quad \lambda_{23}) \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{pmatrix}^{-1} \begin{pmatrix} \lambda_{13} \\ \lambda_{23} \end{pmatrix} \right]^{-1}$$

$$= \left[\lambda_{33} - \frac{\lambda_{13}(\lambda_{13}\lambda_{22} - \lambda_{12}\lambda_{23}) + \lambda_{23}(\lambda_{11}\lambda_{23} - \lambda_{12}\lambda_{13})}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \right]^{-1} \quad (3.4.2)$$

To simplify (3.4.2), only the case $\beta_3=0$ is considered in the search for designs which minimize $\text{var}(\hat{\beta}_3)$. As pointed out in Section 3.1, this gives locally optimal designs. The following theorem shows that optimal designs only need to be found for linear and quadratic parameters which are both positive.

Theorem 3.6. Suppose that $\beta_3=0$, and further that the optimal design for given β_1 and β_2 is comparisons of the pairs (x_i,x_j) a total of n_{x_i,x_j} times, for all pairs. Then

(i) The optimal design for $(-\beta_1)$ and $(-\beta_2)$ is the same as the optimal design for β_1 and β_2,

(ii) The optimal design for $(-\beta_1)$ and β_2 is comparisons of the pairs $(-x_i,-x_j)$ a total of n_{x_i,x_j} times, for all pairs,

(iii) The optimal design for β_1 and $(-\beta_2)$ is the same as in (ii).

Proof. From (3.3.5),

$$\phi_{-\beta_1,-\beta_2}(x_i,x_j) =$$

$$\frac{\exp(-\beta_1(x_i+x_j) - \beta_2(x_i^2+x_j^2))}{(\exp(-\beta_1 x_i - \beta_2 x_i^2) + \exp(-\beta_1 x_j - \beta_2 x_j^2))^2} \cdot \frac{\exp(2\beta_1(x_i+x_j) + 2\beta_2(x_i^2+x_j^2))}{\exp(2\beta_1(x_i+x_j) + 2\beta_2(x_i^2+x_j^2))}$$

$$= \frac{\exp(\beta_1(x_i + x_j) + \beta_2(x_i^2 + x_j^2))}{(\exp(\beta_1 x_i + \beta_2 x_i^2) + \exp(\beta_1 x_j + \beta_2 x_j^2))^2}$$

$$= \phi_{\beta_1,\beta_2}(x_i,x_j) \ . \tag{3.4.3}$$

Notice that since $\beta_3 = 0$, $\pi_i \pi_j / (\pi_i + \pi_j)^2$ from the expression for λ_{ab} given in (2.3.10) is equal to $\phi_{\beta_1,\beta_2}(x_i,x_j)$. Hence by (2.3.10) and (3.4.3), it is clear that λ_{ab} is invariant under changing the sign of both β_1 and β_2, for all a and b. Therefore the variance of $\hat{\beta}_3$ is also invariant under changing the sign of both β_1 and β_2. This implies that the design which minimizes $\text{var}(\hat{\beta}_3)$ for β_1 and β_2 is also the design which minimizes $\text{var}(\hat{\beta}_3)$ for $(-\beta_1)$ and $(-\beta_2)$, thereby proving (i).

To prove (ii), note that from (3.3.5),

$$\phi_{-\beta_1,\beta_2}(x_i,x_j) = \frac{\exp(-\beta_1(x_i+x_j) + \beta_2(x_i^2+x_j^2))}{(\exp(-\beta_1 x_i + \beta_2 x_i^2) + \exp(-\beta_1 x_j + \beta_2 x_j^2))^2}$$

$$= \frac{\exp(\beta_1(-x_i-x_j) + \beta_2(x_i^2+x_j^2))}{(\exp(\beta_1(-x_i) + \beta_2 x_i^2) + \exp(\beta_1(-x_j) + \beta_2 x_j^2))^2}$$

$$= \phi_{\beta_1,\beta_2}(-x_i,-x_j) \ . \tag{3.4.4}$$

Again, by (2.3.10) and (3.4.4), changing β_1 to $(-\beta_1)$ and the pairs (x_i,x_j) to $(-x_i,-x_j)$, for all pairs in the design, will not change λ_{11}, λ_{22}, λ_{33}, nor λ_{13}. It changes only the sign of λ_{12} and λ_{23}. But

notice that in (3.4.2) all terms involving λ_{12} and λ_{23} are of the form $\lambda_{12}\lambda_{23}$, λ_{12}^2, or λ_{23}^2. Hence the negative sign cancels, leaving $\text{var}(\hat{\beta}_3)$ invariant under the change described above. This proves (ii).

To prove (iii), simply apply the results of (i) and (ii).   ÷

Theorem 3.5 proved the equivalence of minimizing $\text{var}(\hat{\beta}_3)$ and El-Helbawy and Bradley's criterion. A Fortran computer program was used to find the optimal design using El-Helbawy and Bradley's methodology for $\beta_1=\beta_2=\beta_3=0$. The levels allowed were x=-1,-.5,0,.5,1. The grid search included all ten pairs formed by these five levels, assuming a symmetric design. The optimal design found was $\{n_{-1,-.5} = n_{.5,1} = .045N, n_{-.5,.5} = .621N, n_{-1,1} = .298N$, all other $n_{ij} = 0\}$. For this design, it can be shown that $\text{var}(\hat{\beta}_3)=16/N$.

It was decided that for unequally spaced levels, employing El-Helbawy and Bradley's methodology would be substantially more laborious than minimizing $\text{var}(\hat{\beta}_3)$. For this reason, the designs subsequently discussed were found by minimizing the expression found in (3.4.2).

Optimal designs were found by a grid search via a Fortran computer program for $\beta_3=0$, $\beta_1,\beta_2=0(.2)1$. The comparisons allowed were $(-1,x_1)$, (x_1,x_2), $(x_2,1)$, and $(-1,1)$, where the optimal value of x_1 and x_2 were also found by the computer. The motivation for the choice of these four pairs is the result discussed in the two preceding paragraphs for the case of all three parameters equal to zero, which indicated that only the pairs $(-1,-.5)$, $(-.5,.5)$, $(.5,1)$, and $(-1,1)$ need to be compared.

The grid procedure is now briefly described. The initial values for x_1 and x_2 were $x_1=\bar{.7}(.1)\bar{.3}$ and $x_2=.3(.1).7$. For each of the 25

combinations of x_1 and x_2, the optimal comparison proportions were found by another grid. The latter grid began by considering different comparison proportions, changing by increments of 0.1. This grid became finer and finer until the proportions were accurate to ±.0004. From this the best combination of x_1 and x_2 out of the initial 25 choices was found, and the entire procedure was repeated for a finer grid about the previous best choice of x_1 and x_2. The entire grid search stopped when x_1 and x_2 were accurate to ±.0037.

Notice that it is possible that the designs found are not absolutely optimal, but rather they may only be local minimums for (3.4.2). This is because only four pairs were considered when it is possible that a five or six pair design, for example, might offer a slight improvement. Also, the grid search restricted x_1 and x_2 rather than allowing them to take on values in the entire interval [-1,1]. The latter point is not believed to be a problem due to the resulting design for $\beta_1=\beta_2=\beta_3=0$ originally found using El-Helbawy and Bradley's methodology, and the fact that the optimal designs currently being discussed all have values of x_1 and x_2 within 0.02 of 0.5.

The optimal designs and the value of $\text{var}(\hat{\beta}_3)$ for N=1 are presented in Table 3.2. The first two columns of the table give the value of β_1 and β_2. Columns 5 through 8 give the comparison proportions for the pairs $(-1,x_1)$, (x_1,x_2), $(x_2,1)$, and $(-1,1)$, respectively, where x_1 and x_2 are given in columns 3 and 4. The last column gives $N \cdot \text{var}(\hat{\beta}_3)$. The number of times a pair is compared is found by multiplying the appropriate comparison proportion by N. Recall that by Theorem 3.6, the optimal design for both parameters negative is the same as the optimal design for $|\beta_1|$ and $|\beta_2|$. If exactly one parameter is negative, then

the optimal design is found by first finding the optimal design for $|\beta_1|$ and $|\beta_2|$, and then taking μ_{-1,x_1} to be the proportion of the total number of comparisons that the pair $(1,-x_1)$ is compared, μ_{x_1,x_2} to be the proportion for $(-x_1,-x_2)$, and so on. To calculate $\text{var}(\hat{\beta}_3)$ for $N>1$, the tabled value of $N \cdot \text{var}(\hat{\beta}_3)$ is divided by N.

Notice from Table 3.2 that for $\beta_1 \geq .6$, the optimal design is very close to the design $\{n_{-1,-.5} = n_{-.5,.5} = n_{.5,1} = .3333N\}$, regardless of the value of β_2. This implies that if it is a priori believed that β_1 is relatively large, then this design will result in a variance of $\hat{\beta}_3$ very close to the tabled minimum. As has been previously mentioned, it is also of interest that the optimal values of x_1 and x_2 are always very close to $-.5$ and $.5$, respectively. Finally, notice that the design for $\beta_1 = \beta_2 = 0$ found in Table 3.2 is not the same as the one found using El-Helbawy and Bradley's methodology, although the value of $\text{var}(\hat{\beta}_3)$ appears to be the same for both, and hence they are equally optimal. However, it was not established that the two designs are mathematically equivalent. Since the two criteria are equivalent, there may be some problem with local minimums of $\text{var}(\hat{\beta}_3)$. It appears that the surface of the expression for $\text{var}(\hat{\beta}_3)$ is relatively flat, and so it is possible that a finer initial grid is necessary to guarantee that the minimum found is in fact the absolute minimum of $\text{var}(\hat{\beta}_3)$.

Table 3.2. Designs which minimize var$(\hat{\beta}_3)$

β_1	β_2	x_1	x_2	μ_{-1,x_1}	μ_{x_1,x_2}	$\mu_{x_2,1}$	$\mu_{-1,1}$	$N\cdot\text{var}(\hat{\beta}_3)$
0	0	-.50	.50	.133	.533	.133	.200	16.0000
.2	0	-.51	.50	.338	.331	.331	0	16.0836
.4	0	-.50	.50	.332	.337	.332	0	16.3225
.6	0	-.50	.50	.330	.341	.330	0	16.7322
.8	0	-.50	.50	.327	.346	.327	0	17.3190
1	0	-.50	.50	.323	.353	.323	0	18.0959
0	.2	-.50	.50	0	.666	0	.333	16.0000
.2	.2	-.50	.51	.333	.329	.338	0	16.1447
.4	.2	-.50	.50	.330	.336	.335	0	16.3835
.6	.2	-.50	.50	.329	.338	.333	0	16.7943
.8	.2	-.50	.50	.326	.346	.328	0	17.3819
1	.2	-.50	.50	.321	.353	.327	0	18.1604
0	.4	-.50	.50	0	.666	0	.333	16.0003
.2	.4	-.50	.50	.105	.561	.103	.231	16.3209
.4	.4	-.50	.50	.330	.333	.337	0	16.5664
.6	.4	-.50	.50	.327	.338	.335	0	16.9799
.8	.4	-.51	.50	.324	.342	.334	0	17.5716
1	.4	-.51	.50	.319	.349	.333	0	18.3554
0	.6	-.50	.50	0	.666	0	.333	16.0005
.2	.6	-.50	.50	.004	.661	.001	.334	16.3207
.4	.6	-.51	.50	.333	.328	.340	0	16.8748
.6	.6	-.51	.50	.327	.333	.340	0	17.2927
.8	.6	-.51	.50	.323	.337	.340	0	17.8907
1	.6	-.51	.49	.317	.346	.337	0	18.6825
0	.8	-.50	.50	0	.666	0	.333	16.0010
.2	.8	-.50	.50	.004	.661	0	.334	16.3225
.4	.8	-.51	.47	.039	.633	.078	.320	17.2619
.6	.8	-.51	.50	.328	.327	.344	0	17.7368
.8	.8	-.51	.50	.323	.331	.346	0	18.3436
1	.8	-.52	.49	.319	.338	.343	0	19.1472
0	1	-.50	.50	0	.666	0	.333	16.0015
.2	1	-.50	.50	0	.664	0	.335	16.3233
.4	1	-.51	.47	.034	.633	0	.333	17.2685
.6	1	-.52	.50	.331	.318	.351	0	18.1804
.8	1	-.52	.50	.324	.325	.351	0	18.9372
1	1	-.52	.49	.318	.331	.351	0	19.7558

CHAPTER 4

DESIGNS FOR FITTING A QUADRATIC MODEL

4.1   Introduction

The present chapter contains optimal designs for fitting a quad-
ratic model under three different conditions, each for three different
optimality criteria.   The D-optimality criterion is considered in Sec-
tion 4.2.   This criterion minimizes the determinant of the asymptotic
variance-covariance matrix.   In the present chapter, this matrix is a
2×2 matrix since there are two maximum likelihood estimators, $\hat{\beta}_1$ and $\hat{\beta}_2$.
From Section 2.3, these estimators are asymptotically normal.   There-
fore, by a result in Box and Lucas (1959), the determinant of the
variance-covariance matrix is proportional to the volume contained with-
in any particular ellipsoidal probability contour for $\underline{\beta} = (\beta_1, \beta_2)$ about
$\underline{\beta}_0$, where $\underline{\beta}_0$ is the true parameter value.   The D-optimal design will
then minimize the volume of any such probability contour.

Section 4.3 contains designs which minimize the average-variance
of $\ln \hat{\pi}_x = \hat{\beta}_1 x + \hat{\beta}_2 x^2$.   The average-variance is simply the variance of
$\ln \hat{\pi}_x$ integrated over the experimental region.   In this case, the exper-
imental region is $x \epsilon [-1,1]$.   The property that these designs have is
obvious.   That is to say, a good estimate of all of the preference prob-
abilities is desirable, and the design which minimizes the integral over
x of $\text{var}(\ln \hat{\pi}_x)$ will give estimates of π_x which are good "on the
average".

The criterion discussed in Section 4.4 is similar to the average-variance criterion. The designs found in this section minimize the maximum of $\text{var}(\ln \hat{\pi}_x)$, where the maximum is taken over the interval $x \in [-1,1]$. These designs are referred to as minimax designs. The disadvantage of this criterion is similar to the disadvantage often attributable to the minimax rules in the decision theory framework. That is, minimax designs may be slightly "pessimistic" in the sense that they minimize the largest value that the variance can be. This minimization is good, but at the same time the variance of $\ln \hat{\pi}_x$ may be relatively large for most values of x in the interval $[-1,1]$, hence causing the minimax design to be unfavorable.

Finally, Section 4.5 presents some design recommendations and conclusions. A number of designs are compared to determine which design does best for a wide range of parameter values, using primarily the D-optimality criterion. Additionally, an examination is performed on how well these designs protect against the bias present when the true model is cubic.

Recall that in general, the problem of finding optimal designs is complex. It involves the determination of the number of levels to be included in the experimental design, and also which particular levels are to be chosen. Additionally, the number of times each pair is to be compared must be determined. In Chapter 3 it was proven that the design which minimizes $\text{var}(\hat{\beta}_2)$ must be of the form $\{n_{-1,-x} = n_{x,1} = N/2\}$, thereby implying that the optimal number of levels is four, and only two pairs are to be compared an equal number of times. All that remained to be resolved was the determination of x for different values of $|\beta_1|$. In the present situation, a similar simplification of the

general problem does not appear to be possible, except by imposing some restrictions, such as letting $\beta_2=0$. For this reason, three conditions are considered under which optimal designs are found. This essentially results in designs which are optimal for a proper subclass of the class of all designs. However, as will become apparent in the remainder of the chapter, there is not much difference in the degree of optimality among the three conditions presently considered. Hence it is conjectured that the degree of optimality of the designs found in the final and most general subsection of Sections 4.2-4.4 is very close to if not equal to the best that can possibly be achieved. The three conditions are now briefly introduced.

The first condition considered in each of Sections 4.2-4.4 is that the levels are restricted to be x=-1,0,1. This greatly simplifies the general problem to one with only two unknowns, namely the comparison proportions, $\mu_{-1,0}$ and $\mu_{0,1}$. The value of $\mu_{-1,1}$ is found by the relationship $\mu_{-1,0} + \mu_{0,1} + \mu_{-1,1} = 1$.

The second condition discussed in each of the following three sections is to let $\beta_2=0$ in the enumeration of the variances and covariance. In this situation the optimal designs are proved to be symmetric, and hence the design problem is simplified to a choice from the class of symmetric designs. As was pointed out in Section 2.3, previous papers on the design of paired comparison experiments have always relied upon setting all parameters equal to zero. Presently, the linear parameter is allowed to vary. The three criteria are a continuous function of the parameters because the variance-covariance matrix is continuous in the parameters. Therefore, the designs found for the second condition

are locally optimal in the sense that they are close to optimal for values of β_2 which are near zero.

The third condition in each of the following three sections is a consideration of designs whose comparisons are of the form $(-1,x_1)$, $(x_2,1)$, and $(-1,1)$. The optimal choice of the levels x_1 and x_2 are found in addition to the three comparison proportions. This is a generalization of the first condition since setting $x_1=x_2=0$ results in the first condition.

Figure 4.1 presents a synopsis of the procedures used in Sections 4.2-4.4 to find the optimal designs for each of the three optimality criteria and the three conditions introduced in the present section. The first column presents a brief description of the three conditions, and hence the three rows in the figure correspond to the three conditions. Columns 2-4 contain a brief description of the methodology used for the three criteria.

4.2   D-optimal Designs

The D-optimal design is defined to be the design which minimizes the determinant of the asymptotic variance-covariance matrix. Letting Σ denote the variance-covariance matrix, this is equivalent to maximizing $|\Sigma^{-1}|$. In the case of a quadratic model, the D-optimal design is the design which maximizes D, where

$$D = |\Sigma^{-1}| = \lambda_{11}\lambda_{22} - \lambda_{12}^2 , \qquad (4.2.1)$$

and where λ_{11}, λ_{22}, and λ_{12} are given in (3.3.2)-(3.3.4). Because the optimal design depends on the true parameter values, designs are found

| Condition | - - - - - Three optimality criteria - - - - - | | |
	D-optimal	Average-variance	Minimax
1 $\beta_1,\beta_2 = 0(.1)1,2,3.$ Optimal comparison proportions for $(-1,0)$, $(0,1)$, and $(-1,1)$ are found.	Analytical derivation of optimal proprotions $(\partial D/\partial \mu_{ij} = 0)$.	Grid search	Grid search
2 $\beta_1 = 0(.02)1,2(1)5, \beta_2 = 0.$ Optimal design is symmetric about 0 and all comparisons involve $x = \pm 1$. Best x for $N/2$ comparisons of each of $(-1,-x)$ and $(x,1)$ is found.	Analytical derivation of x $(\partial D/\partial x = 0)$. Algebraically shown to give local optimum.	Analytical derivation of x $(\partial V/\partial x = 0)$. Shown to give local optimum.	Analytical derivation of x $(\partial M/\partial x = 0)$. Shown to give local optimum.
3 $\beta_1,\beta_2 = 0(.1)1,2,3.$ Best x_1 and x_2 and optimal comparison proportions for $(-1,x_1)$, $(x_2,1)$, and $(-1,1)$ are found.	Grid search	Grid search	Grid search

Figure 4.1. Procedure summary

for various choices of the parameters. The three conditions intro-
duced in Section 4.1 are now discussed, beginning with Condition 1.

Condition 1

The levels are restricted to be x=-1,0,1. Because of this restric-
tion, and by substituting μ_{ij} for n_{ij} in (3.3.2)-(3.3.4), the expression
for λ_{11}, λ_{12}, and λ_{22} can be simplified to

$$\lambda_{11} = \mu_{-1,0}\phi_{-1,0} + \mu_{0,1}\phi_{0,1} + 4\mu_{-1,1}\phi_{-1,1} \, ,$$

$$\lambda_{12} = \mu_{0,1}\phi_{0,1} - \mu_{-1,0}\phi_{-1,0} \, , \qquad (4.2.2)$$

$$\lambda_{22} = \mu_{-1,0}\phi_{-1,0} + \mu_{0,1}\phi_{0,1} \, .$$

The substitution of μ_{ij} for n_{ij} is made throughout the present chapter.
This essentially sets N=1 since $\mu_{ij} = \lim_{N \to \infty}(n_{ij}/N)$.

From (4.2.1) and (4.2.2), the expression for D is

$$D = 4\mu_{-1,0}\mu_{0,1}\phi_{-1,0}\phi_{0,1} +$$

$$4(1-\mu_{-1,0} - \mu_{0,1})(\mu_{-1,0}\phi_{-1,0} + \mu_{0,1}\phi_{0,1})\phi_{-1,1} \, , \qquad (4.2.3)$$

where ϕ_{ij} is defined in (3.3.5). The partial derivatives of D with
respect to $\mu_{-1,0}$ and $\mu_{0,1}$ are, respectively,

$$\frac{\partial D}{\partial \mu_{-1,0}} = 4\mu_{0,1}\phi_{-1,0}\phi_{0,1} - 4\phi_{-1,1}(\mu_{-1,0}\phi_{-1,0} + \mu_{0,1}\phi_{0,1}) +$$

$$(4.2.4)$$

$$4(1 - \mu_{-1,0} - \mu_{0,1})\phi_{-1,1}\phi_{-1,0} \, ,$$

$$\frac{\partial D}{\partial \mu_{0,1}} = 4\mu_{-1,0}\phi_{-1,0}\phi_{0,1} - 4\phi_{-1,1}(\mu_{-1,0}\phi_{-1,0} + \mu_{0,1}\phi_{0,1}) +$$

$$4(1 - \mu_{-1,0} - \mu_{0,1})\phi_{-1,1}\phi_{0,1} \ . \tag{4.2.4}$$

Setting the derivatives in (4.2.4) equal to zero results in

$$\begin{bmatrix} \mu_{-1,0}(2\phi_{-1,0}\phi_{-1,1}) + \\ \mu_{0,1}(-\phi_{-1,0}\phi_{0,1} + \phi_{0,1}\phi_{-1,1} + \phi_{-1,0}\phi_{-1,1}) \end{bmatrix} = \phi_{-1,0}\phi_{-1,1} \ ,$$

$$\begin{bmatrix} \mu_{-1,0}(-\phi_{-1,0}\phi_{0,1} + \phi_{-1,0}\phi_{-1,1} + \phi_{0,1}\phi_{-1,1}) \\ + \mu_{0,1}(2\phi_{0,1}\phi_{-1,1}) \end{bmatrix} = \phi_{0,1}\phi_{-1,1} \ . \tag{4.2.5}$$

In matrix notation, (4.2.5) is equivalent to

$$A \begin{pmatrix} \mu_{-1,0} \\ \mu_{0,1} \end{pmatrix} = \begin{pmatrix} \phi_{-1,0}\phi_{-1,1} \\ \phi_{0,1}\phi_{-1,1} \end{pmatrix} \ , \tag{4.2.6}$$

implying

$$\begin{pmatrix} \mu_{-1,0} \\ \mu_{0,1} \end{pmatrix} = A^{-1} \begin{pmatrix} \phi_{-1,0}\phi_{-1,1} \\ \phi_{0,1}\phi_{-1,1} \end{pmatrix} \ , \tag{4.2.7}$$

where

$$A = \begin{pmatrix} 2\phi_{-1,0}\phi_{-1,1} & -\phi_{-1,0}\phi_{0,1} + \phi_{0,1}\phi_{-1,1} + \phi_{-1,0}\phi_{-1,1} \\ -\phi_{-1,0}\phi_{0,1} + \phi_{-1,0}\phi_{-1,1} + \phi_{0,1}\phi_{-1,1} & 2\phi_{0,1}\phi_{-1,1} \end{pmatrix} .$$

For the solution of (4.2.7) to be a maximum for the expression in (4.2.3), it needs to be shown that the matrix of second order partial

derivatives is negative definite (see Kaplan, page 128). By (4.2.4),

this matrix is

$$
X \;=\; \begin{pmatrix} \partial^2 D/\partial \mu_{-1,0}{}^2 & \partial^2 D/\partial \mu_{-1,0}\partial \mu_{0,1} \\[2ex] \partial^2 D/\partial \mu_{-1,0}\partial \mu_{0,1} & \partial^2 D/\partial \mu_{0,1}{}^2 \end{pmatrix} \;=\; -4\,A\,, \qquad (4.2.8)
$$

where, from (3.3.5),

$$
\phi_{-1,1} \;=\; 1/(\,e^{-\beta_1} + e^{\beta_1}\,)^2\,,
$$

$$
\phi_{-1,0} \;=\; e^{-\beta_1 + \beta_2}/(\,e^{-\beta_1 + \beta_2} + 1\,)^2\,, \qquad (4.2.9)
$$

$$
\phi_{0,1} \;=\; e^{\beta_1 + \beta_2}/(\,1 + e^{\beta_1 + \beta_2}\,)^2\,.
$$

It is known from elementary matrix algebra that the determinant of

a square matrix is equal to the product of its eigenvalues. So in this

case, $|X| = e_1 e_2$, where e_1 and e_2 are the eigenvalues of X. If $|X| > 0$,

then either both eigenvalues are positive or both are negative, or

equivalently, X is either positive definite or negative definite. But

it is clear from (4.2.8) that

$$
(\,1 \quad 0\,)\; X \begin{pmatrix} 1 \\ 0 \end{pmatrix} \;=\; -8\phi_{-1,1}\phi_{-1,0} \;<\; 0\,, \qquad (4.2.10)
$$

where $\phi_{-1,1}$ and $\phi_{-1,0}$ are both positive and are given in (4.2.9). The

inequality in (4.2.10) implies that X is negative definite if $|X| > 0$.

However, it will be shown by (4.2.12) that the determinant is not nec-

essarily positive for all values of β_1 and β_2.

To simplify the notation, we write

$$X = \begin{pmatrix} -2a & -(a + c - b) \\ -(a + c - b) & -2c \end{pmatrix},$$

where

$$a = \phi_{-1,1}\phi_{-1,0} = e^{-\beta_1 + \beta_2} / \left[(e^{-\beta_1} + e^{\beta_1})(e^{-\beta_1 + \beta_2} + 1) \right]^2,$$

$$b = \phi_{-1,0}\phi_{0,1} = e^{2\beta_2} / \left[(e^{-\beta_1 + \beta_2} + 1)(1 + e^{\beta_1 + \beta_2}) \right]^2,$$

$$c = \phi_{-1,1}\phi_{0,1} = e^{\beta_1 + \beta_2} / \left[(e^{-\beta_1} + e^{\beta_1})(1 + e^{\beta_1 + \beta_2}) \right]^2.$$

Then the determinant of X can be written as

$$|X| = 4ac - (a + c - b)^2$$

$$= 2ac + 2ab + 2bc - a^2 - b^2 - c^2. \qquad (4.2.11)$$

After resubstituting for a, b, and c, and after some algebra in (4.2.11), an expression for $|X|$ is found by

$$\frac{|X|}{\xi} = 2 \left[\begin{array}{l} e^{4\beta_2} + 4e^{-\beta_1 + 3\beta_2} + 1 + 4e^{\beta_1 + \beta_2} + 3e^{-\beta_1 + \beta_2} \\[4pt] + 3e^{\beta_1 + 3\beta_2} + e^{-3\beta_1 + \beta_2} + 3e^{-2\beta_1 + 2\beta_2} + e^{3\beta_1 + 3\beta_2} \\[4pt] + 3e^{2\beta_1 + 2\beta_2} + 3e^{2\beta_2} + e^{-3\beta_1 + 3\beta_2} + e^{3\beta_1 + \beta_2} \end{array} \right]$$

$$\qquad (4.2.12)$$

$$- \left[\begin{array}{l} e^{-2\beta_1} + 2e^{-\beta_1} + e^{2\beta_1 + 4\beta_2} + e^{-4\beta_1 + 2\beta_2} \\[4pt] + e^{4\beta_1 + 2\beta_2} + e^{-2\beta_1 + 4\beta_2} + e^{2\beta_1} + 2e^{\beta_1 + 2\beta_2} \end{array} \right],$$

where

$$\xi = 4e^{2\beta_2} / \left[(e^{-\beta_1} + e^{\beta_1})(e^{-\beta_1 + \beta_2} + 1)(1 + e^{\beta_1 + \beta_2}) \right]^2 . \quad (4.2.13)$$

Since $\xi > 0$ for all β_1 and β_2, $|X| > 0$ if and only if the right-hand side of (4.2.12) is positive. However, it appears that for large enough values of β_1 and β_2, some of the terms with a negative sign will overwhelm the sum of all of the terms with a positive sign, and hence the determinant could be negative. For example, β_1 and β_2 can be made large enough such that the term $\exp(4\beta_1 + 2\beta_2)$ in (4.2.12) is larger than the sum of all the terms with positive signs, thereby making the expression in (4.2.12) negative. It will be seen shortly that this is in fact the case.

An APL computer program was utilized to solve equation (4.2.7) for various values of β_1 and β_2. The determinant, $|X|$, was also calculated and found to be positive for all values of β_1 and β_2 when the solution fell inside the region of allowable values for the comparison proportions, $\mu_{-1,0}$, $\mu_{0,1}$, and $\mu_{-1,1}$. That is, $\mu_{ij} \geq 0$, for all i and j, and $\mu_{-1,0} + \mu_{0,1} + \mu_{-1,1} = 1$. Figure 4.2 depicts the values that $\mu_{-1,0}$ and $\mu_{0,1}$ can have. They must lie inside or on the solid triangle. The dotted triangle in the figure depicts the region that the D-optimal designs are exclusively located.

When the solution to (4.2.7) falls outside of the solid triangle in Figure 4.2, the optimal design must lie somewhere on the solid triangle. This can be proven by contradiction since D is a quadratic function of two unknowns, and hence it can only have one stationary point. For a point to be on the triangle, one of the following must be true: $\mu_{-1,0} = 0$, $\mu_{0,1} = 0$, or $\mu_{-1,0} + \mu_{0,1} = 1$.
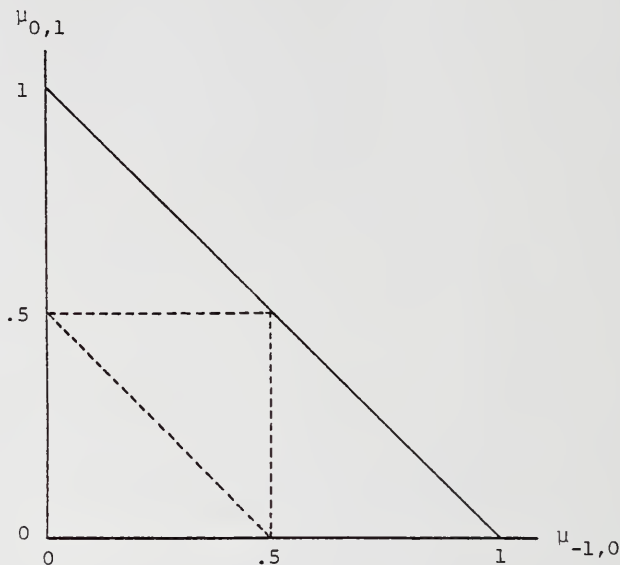
Figure 4.2. Depiction of possible designs for Condition 1

If $\mu_{-1,0} = 0$, then from (4.2.3),

$$D = 4\phi_{0,1}\phi_{-1,1}(\mu_{0,1} - \mu_{0,1}^2) \, ,$$

and $\partial D/\partial \mu_{0,1} = 0$ implies $\mu_{0,1} = 0.5$. This clearly maximizes D because of the negative coefficient on the quadratic term. Therefore if the optimal design lies on the line segment $\mu_{-1,0} = 0$, then it must be $\{\mu_{0,1} = \mu_{-1,1} = 0.5\}$. If $\mu_{0,1} = 0$, then

$$D = 4\phi_{-1,1}\phi_{-1,0}(\mu_{-1,0} - \mu_{-1,0}^2) \, ,$$

and it can similarly be shown that the optimal design is the midpoint

of the line segment, namely $\{\mu_{0,1} = 0, \ \mu_{-1,0} = \mu_{-1,1} = 0.5\}$. Finally, if $\mu_{-1,0} + \mu_{0,1} = 1$, then

$$D = 4\mu_{-1,0}\mu_{0,1}\phi_{-1,0}\phi_{0,1}$$

$$= 4\phi_{-1,0}\phi_{0,1}\mu_{-1,0}(1 - \mu_{-1,0}) \ .$$

Again, it can similarly be shown that D is maximized when $\mu_{-1,0} = \mu_{0,1} = 0.5$. Therefore, if the D-optimal design is on the solid tri-angle, it must be one of these three designs.

When the solution to (4.2.7) was outside of the solid triangle in Figure 4.2, D was calculated for the three designs presented in the preceding paragraph, and the one with the largest value of D was taken as D-optimal. Otherwise the D-optimal design was found directly from (4.2.7). The determinant of X was also calculated. It was occasion-ally negative as conjectured earlier, but only when the solution was outside of the region, at which time the sign of $|X|$ is irrelevant.

The "restricted" D-optimal designs are presented in Table 4.1 for all combinations of the parameter values $\beta_1 = 0(.1)1,2,3$ and $\beta_2 = 0(.1)1,2,3$. Recall that $\mu_{-1,0}, \ \mu_{0,1},$ and $\mu_{-1,1}$ represent the proportion of the total available comparisons to be allocated to the pairs $(-1,0)$, $(0,1)$, and $(-1,1)$, respectively. The number of times one of these three pairs should be compared is then found by multiplying the corresponding μ_{ij} by N. The values of D are also given for the purpose of comparing the designs with the designs found in the next two subsections. Addition-ally, the relative efficiencies are given for the "restricted" D-opti-mal designs presented in Table 4.1 relative to the D-optimal designs

found for Condition 3 in Table 4.3. The relative efficiency is the square root of the ratio of the corresponding values of D, since D is a multiple of N^2.

Recall from Section 3.1 that a value of 1 for the parameter β_1 is "large". This is basically true for any parameter since x is in the interval $[-1,1]$. For this reason, it is assumed that rarely will β_1 or β_2 be larger than 1, and hence parameter values of less than or equal to 1 are of most interest. Parameter values of 2 and 3 are included primarily to observe the change in the designs for large parameter values.

Table 4.1 only gives positive parameter values. The following theorem shows that the table can also be used to find optimal designs for negative values of β_1 or β_2. This theorem is similar to Theorem 3.6.

Theorem 4.1. Suppose that the D-optimal design for given β_1 and β_2 is a comparison of x_i with x_j a total of n_{x_i,x_j} times, for all pairs in the design. Then

(i) The D-optimal design for $(-\beta_1)$ and $(-\beta_2)$ is the same as for β_1 and β_2,

(ii) The D-optimal design for $(-\beta_1)$ and β_2 is a comparison of $(-x_i)$ with $(-x_j)$ a total of n_{x_i,x_j} times, for all pairs in the design,

(iii) The D-optimal design for β_1 and $(-\beta_2)$ is the same as in (ii).

Proof. From (3.4.3),

$$\phi_{-\beta_1,-\beta_2}(x_i,x_j) \;=\; \phi_{\beta_1,\beta_2}(x_i,x_j) \ ,$$

where $\phi_{\beta_1,\beta_2}(x_i,x_j)$ is given in (3.3.5). From (3.3.2)-(3.3.4) and the above result, it is clear that λ_{11}, λ_{22}, and λ_{12} are invariant under changing the sign of both β_1 and β_2. Therefore, since $D = \lambda_{11}\lambda_{22}-\lambda_{12}^2$, D is invariant under changing the sign of both β_1 and β_2, and so (i) is proved.

To prove (ii), note that by (3.4.4),

$$\phi_{-\beta_1,\beta_2}(x_i,x_j) = \phi_{\beta_1,\beta_2}(-x_i,-x_j) .$$

Again, from (3.3.2)-(3.3.4) and the above, changing β_1 to $(-\beta_1)$ and the pairs (x_i,x_j) to $(-x_i,-x_j)$ will not change λ_{11} nor λ_{22}. It changes only the sign of λ_{12}, and hence does not change λ_{12}^2. Therefore (ii) is proved.

To prove (iii), simply apply the results of (i) and (ii).   +

So, as in Section 3.3, the optimal design for both parameters negative is the same as the design for $|\beta_1|$ and $|\beta_2|$. When exactly one parameter is negative, the column headed $\mu_{-1,0}$ becomes $\mu_{0,1}$, and vice versa.

Notice from Table 4.1 that for small parameter values, the optimal design is inside the solid triangle in Figure 4.2, signifying all three pairs are to be compared. The optimal design approaches the boundary as the parameters increase. Also, in all cases the optimal design was inside or on the dotted triangle in Figure 4.2, signifying that no pair should be compared more than half of the time.

Notice also from Table 4.1 that most of the relative efficiencies are approximately 0.95 when β_1 and β_2 are both less than or equal to 1.

This indicates that the increase in efficiency for the generalization considered under Condition 3 is slight. If an experimenter warrants that a minimum number of levels should be used, then the advantage of the designs found presently under Condition 1 is that they have only three levels as opposed to four levels for the other two conditions. The fact that they are also about 95% efficient allows the present designs to be useful in this situation. However, if there is no signi-ficant advantage to only having three levels, then perhaps the designs found under Conditions 2 and 3 are more useful since they are slightly more efficient.

Condition 2

In this subsection, locally D-optimal designs are found analytical-ly for $\beta_2 = 0$. The designs are local as described in Section 4.1. That is, by the continuity of the criterion, the optimal designs are close to optimal for β_2 close to zero. It turns out that they are also local in the sense that the designs found under Condition 3 have slightly larger values of D for some values of β_1. Since the criterion is a maximization of D, the designs under Condition 3 are therefore occa-sionally slightly better.

By the same argument that was used in the proof of Theorem 3.1, the D-optimal design must be a symmetric design when $\beta_2 = 0$, which implies $\lambda_{12}^2 = 0$. So the search for a D-optimal design is reduced to a search for a symmetric design which maximizes $D = \lambda_{11}\lambda_{22}$. Also, the proof of Theorem 3.2 can be presently implemented to show that every pair in the D-optimal design must include either x=-1 or x=1.

From (3.3.7), ϕ_{ij} is written as

$$\phi_{ij} = (\, 4\cosh^2(\beta_1(x_i - x_j)/2) \,)^{-1} \, ,$$

where

$$\cosh(u) = \frac{1}{2} (e^u + e^{-u}) \, ,$$

$$\sinh(u) = \frac{1}{2} (e^u - e^{-u}) \, .$$

The maximization of D is equivalent to the maximization of the following redefined D:

$$D = 4\lambda_{11}\lambda_{22} = f(\underline{x})g(\underline{x}) \, , \qquad (4.2.14)$$

where

$$f(\underline{x}) = \sum_{i=1}^{N/2} \frac{(1 - x_i)^2}{\cosh^2(\beta_1(x_i - 1)/2)} \, , \qquad (4.2.15)$$

$$g(\underline{x}) = \sum_{i=1}^{N/2} \frac{(1 - x_i)^2(1 + x_i)^2}{\cosh^2(\beta_1(x_i - 1)/2)} \, , \qquad (4.2.16)$$

$$\underline{x}' = (x_1, x_2, \ldots, x_{N/2}) \, .$$

Two useful, easily verified facts are

$$\frac{\partial \cosh(\beta_1(x_i - 1)/2)}{\partial x_i} = \frac{1}{2} \beta_1 \sinh(\beta_1(x_i - 1)/2) \, ,$$

$$(4.2.17)$$

$$\frac{\partial \sinh(\beta_1(x_i - 1)/2)}{\partial x_i} = \frac{1}{2} \beta_1 \cosh(\beta_1(x_i - 1)/2) \, .$$

The partial derivative of D with respect to x_i is then

$$\frac{\partial D}{\partial x_i} = \frac{\partial f(\underline{x})}{\partial x_i} g(\underline{x}) + f(\underline{x}) \frac{\partial g(\underline{x})}{\partial x_i} , \qquad (4.2.18)$$

where

$$\frac{\partial f(\underline{x})}{\partial x_i} = \frac{-2(1 - x_i)}{\cosh^2(\beta_1(x_i - 1)/2)} - \frac{(1 - x_i)^2 \beta_1 \sinh(\beta_1(x_i - 1)/2)}{\cosh^3(\beta_1(x_i - 1)/2)} , \qquad (4.2.19)$$

$$\frac{\partial g(\underline{x})}{\partial x_i} = \frac{4x_i^3 - 4x_i}{\cosh^2(\beta_1(x_i - 1)/2)} - \frac{(1 - x_i^2)^2 \beta_1 \sinh(\beta_1(x_i - 1)/2)}{\cosh^3(\beta_1(x_i - 1)/2)} , \qquad (4.2.20)$$

$$i = 1, \ldots, N/2.$$

Because of the complexity of handling these N/2 equations of N/2 unknowns, and because it appeared to be reasonable, the expression in (4.2.18) was evaluated at the N/2-dimensional point $x_i = x$, $i=1,\ldots,N/2$, set equal to zero and solved. This produced a design of two equally weighted comparisons, namely $(-1,-x)$ and $(x,1)$, for some x. The mechanics of the procedure follows.

The partial derivative of D with respect to x_i, evaluated at $x_i = x$, for all i, is

$$\left. \frac{\partial D}{\partial x_i} \right|_{x_i = x} = \left\{ \left[\frac{-2(1 - x)}{\cosh^2} - \frac{\beta_1(1 - x)^2 \sinh}{\cosh^3} \right] \left[\frac{\frac{N}{2}(1 - x^2)^2}{\cosh^2} \right] \right.$$
$$\left. + \left[\frac{\frac{N}{2}(1 - x)^2}{\cosh^2} \right] \left[\frac{-4x(1 - x^2)}{\cosh^2} - \frac{\beta_1(1 - x^2)^2 \sinh}{\cosh^3} \right] \right\} , \qquad (4.2.21)$$

where cosh and sinh are shorthand for $\cosh(\beta_1(x - 1)/2)$ and

$\sinh(\beta_1(x - 1)/2)$, respectively. By setting the expression in (4.2.21) equal to zero, multiplying it by $\cosh^5/(N(1 - x)^3(1 + x))$, and combining terms, we have

$$(1 + 3x)\cosh + \beta_1(1 - x^2)\sinh = 0. \tag{4.2.22}$$

Because of the transcendental nature of this equation, x can not be expressed as a closed form function of β_1. Consequently, an APL program was written to find the solution to (4.2.22) for various choices of β_1. These designs are presented in Table 4.2 and are discussed at the end of the present subsection.

In order for the solution to (4.2.22) to be a local maximum for D, it must be shown that

$$(\cos\alpha_1, \ldots, \cos\alpha_{N/2}) \left(\left. \frac{\partial^2 D}{\partial x_i \partial x_j} \right|_{x_i = x} \right) \begin{pmatrix} \cos\alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \cos\alpha_{N/2} \end{pmatrix} < 0$$

for any angles α_i, such that $0 \le \alpha_i \le 2\pi$, and

$$\sum_{i=1}^{N/2} \cos^2\alpha_i = 1$$

(see Kaplan, page 128). This is equivalent to showing that the matrix X is negative definite, where

$$X = \left(\left. \frac{\partial^2 D}{\partial x_i \partial x_j} \right|_{x_i = x} \right)_{N/2 \times N/2}. \tag{4.2.23}$$

From (4.2.18), the second order partial derivatives are

$$\frac{\partial^2 D}{\partial x_i^2} = \frac{\partial^2 f(\underline{x})}{\partial x_i^2} g(\underline{x}) + 2\frac{\partial f(\underline{x})}{\partial x_i} \frac{\partial g(\underline{x})}{\partial x_i} + f(\underline{x})\frac{\partial^2 g(\underline{x})}{\partial x_i^2} , \quad i=1,\ldots,N/2,$$
(4.2.24)

$$\frac{\partial^2 D}{\partial x_i \partial x_j} = \left[\frac{\partial^2 f(\underline{x})}{\partial x_i \partial x_j} g(\underline{x}) + \frac{\partial f(\underline{x})}{\partial x_i} \frac{\partial g(\underline{x})}{\partial x_j}\right] + \left[\frac{\partial f(\underline{x})}{\partial x_j} \frac{\partial g(\underline{x})}{\partial x_i} + f(\underline{x})\frac{\partial^2 g(\underline{x})}{\partial x_i \partial x_j}\right]$$

$$= \frac{\partial f(\underline{x})}{\partial x_i} \frac{\partial g(\underline{x})}{\partial x_j} + \frac{\partial f(\underline{x})}{\partial x_j} \frac{\partial g(\underline{x})}{\partial x_i} , \quad i \neq j, \ i,j=1,\ldots,N/2, \quad (4.2.25)$$

where, from (4.2.19) and (4.2.20),

$$\frac{\partial^2 f(\underline{x})}{\partial x_i^2} = \frac{2 - \frac{1}{2}(1-x_i)^2\beta_1^2}{\cosh^2} + \frac{4(1-x_i)\beta_1\sinh}{\cosh^3} - \frac{\frac{3}{2}(1-x_i)^2\beta_1^2\sinh^2}{\cosh^4} ,$$
(4.2.26)

$$\frac{\partial^2 g(\underline{x})}{\partial x_i^2} = \left[\frac{4(3x_i^2-1) - \frac{1}{2}(1-x_i^2)^2\beta_1^2}{\cosh^2} - \frac{8(x_i^3-x_i)\beta_1\sinh}{\cosh^3} + \frac{\frac{3}{2}(1-x_i^2)^2\beta_1^2\sinh^2}{\cosh^4}\right] \quad i=1,\ldots,N/2.$$
(4.2.27)

Evaluating the second order partial derivatives at the point $x_i=x$, for all i, (4.2.23) can be written as

$$X = \begin{pmatrix} a & b & . & . & . & b \\ b & a & & & & . \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & a & b \\ b & . & . & . & b & a \end{pmatrix} ,$$
(4.2.28)

where

$$a = \left. \frac{\partial^2 D}{\partial x_i^2} \right|_{x_i = x}$$

$$= \left[\frac{\partial^2 f(\underline{x})}{\partial x_i^2} g(\underline{x}) + 2 \frac{\partial f(\underline{x})}{\partial x_i} \frac{\partial g(\underline{x})}{\partial x_i} + f(\underline{x}) \frac{\partial^2 g(\underline{x})}{\partial x_i^2} \right] \Bigg|_{x_i = x} \qquad (4.2.29)$$

and

$$b = \left. \frac{\partial^2 D}{\partial x_i \partial x_j} \right|_{x_i = x} \qquad \text{(any } i \neq j)$$

$$= 2 \left[\frac{\partial f(\underline{x})}{\partial x_i} \frac{\partial g(\underline{x})}{\partial x_i} \right] \Bigg|_{x_i = x} . \qquad (4.2.30)$$

The complete expressions for a and b can be found from (4.2.29) and (4.2.30), in addition to (4.2.15), (4.2.16), (4.2.19), and (4.2.20), by substituting x for x_i.

The matrix X is of a special form as indicated by (4.2.28). It can be written as

$$X = D + \alpha \underline{y} \underline{z}' ,$$

where $D = (a - b) I_{N/2}$, $\alpha = b$, and $\underline{y}' = \underline{z}' = \underline{1}' = (1, \ldots, 1)$. From Graybill (page 187, Theorem 8.5.3), the characteristic equation is

$$\left(d + \alpha \sum_{i=1}^{N/2} y_i z_i - \lambda \right) (d - \lambda)^{\frac{N}{2} - 1} = 0 .$$

So $(\frac{N}{2} - 1)$ of the eigenvalues are d=a-b, and one eigenvalue is $d + \alpha \sum_i y_i z_i = (a - b) + \frac{N}{2}b$. As discussed in the previous subsection, the matrix X is negative definite if and only if all of the eigenvalues are negative. An analytic proof that the two eigenvalues are negative follows.

Consider first the eigenvalue (a-b), where from (4.2.29) and (4.2.30),

$$a-b = \left[\frac{\partial^2 f(\underline{x})}{\partial x_i^2} g(\underline{x}) + f(\underline{x}) \frac{\partial^2 g(\underline{x})}{\partial x_i^2}\right]\bigg|_{x_i = x}$$

$$= \left\{ \frac{\left[\cosh^2(2 - \frac{1}{2}(1-x)^2\beta_1^2) + \sinh\cdot\cosh\cdot(4(1-x)\beta_1) + \sinh^2(-\frac{3}{2}(1-x)^2\beta_1^2)\right]}{\cosh^4} \left[\frac{\frac{N}{2}(1-x^2)^2}{\cosh^2}\right] + \left[\frac{\frac{N}{2}(1-x)^2}{\cosh^2}\right]\left[\frac{\cosh^2(4(3x^2-1) - \frac{1}{2}(1-x^2)\beta_1^2 + \sinh\cdot\cosh\cdot(8x(1-x^2)\beta_1) + \sinh^2(\frac{3}{2}(1-x^2)^2\beta_1^2)}{\cosh^4}\right] \right\} .$$

$$(4.2.31)$$

It needs to be shown that (a-b) is negative at the stationary point x, which is the solution to (4.2.22). This relationship between x and β_1 found in (4.2.22) is used in the above expression of (a-b) by substituting for sinh,

$$\sinh = \frac{-(1 + 3x)\cosh}{\beta_1(1 - x^2)} .$$

$$(4.2.32)$$

Upon substituting (4.2.32) into (4.2.31), the eigenvalue (a-b) at the stationary point, x, is

$$
a-b = \left\{
\begin{aligned}
&\left[(2 - \tfrac{1}{2}(1-x)^2\beta_1^2) + \frac{-4(1+3x)}{(1+x)} - \frac{3(1+3x)^2}{2(1+x)^2}\right]\left[\frac{\tfrac{N}{2}(1-x)^2(1+x)^2}{\cosh^4}\right] \\
&+ \left[(4(3x^2-1) - \tfrac{1}{2}(1-x^2)\beta_1^2) - 8x(1+3x) + \tfrac{3}{2}(1+3x)^2\right]\left[\frac{\tfrac{N}{2}(1-x)^2}{\cosh^4}\right]
\end{aligned}
\right\} .
$$

$$(4.2.33)$$

After factoring out $\tfrac{N}{2}(1-x)^2/\cosh^4$ and combining terms,

$$ a-b \;<\; 0 $$

if and only if (4.2.34)

$$ \beta_1^2(-\tfrac{1}{2}(x^2 - 1)(x^2 - 2)) + (-22x^2 - 20x - 6) \;<\; 0 . $$

The first term in (4.2.34) has roots $x=\pm1$ and $x=\pm\sqrt{2}$, and is nonpositive for $x\in[-1,1]$. The roots of the second term are imaginary, and clearly the function is negative at x=0. Hence it is negative for all x. The result a-b < 0 then follows from (4.2.34).

The second eigenvalue is $(a-b) + \tfrac{N}{2}b$. From (4.2.19), (4.2.20), and (4.2.30),

$$
\begin{aligned}
b &= \frac{2(1-x)}{\cosh^6} (-2\cosh - (1-x)\beta_1\sinh)(-4x(1+x)\cosh - (1-x)(1+x)^2\beta_1\sinh) \\[4pt]
&= \frac{2(1-x)}{\cosh^6} \left[
\begin{aligned}
&8x(1+x)\cosh^2 + 2(1-x)(1+x)^2\beta_1\cosh\cdot\sinh \\
&+ 4x(1-x^2)\beta_1\cosh\cdot\sinh + (1-x^2)^2\beta_1^2\sinh^2
\end{aligned}
\right] \\[4pt]
&= \frac{2(1-x)}{\cosh^4} (-x^2 + 2x - 1) ,
\end{aligned}
$$

$$(4.2.35)$$

the last step produced by again using the relationship in (4.2.32) and combining terms. Then

$$\frac{N}{2} b = \frac{\frac{N}{2}(1-x)^2}{\cosh^4}(2x-2) . \qquad (4.2.36)$$

By factoring $\frac{N}{2}(1-x)^2/\cosh^4$ out of the second eigenvalue, it is clear from (4.2.34) and (4.2.36) that

$$(a - b) + \frac{N}{2} b < 0$$

if and only if

$$\beta_1^2(-\frac{1}{2}(x^2-1)(x^2-2)) + (-22x^2-20x-6) + (2x-2) < 0$$

if and only if

$$\beta_1^2(-\frac{1}{2}(x^2-1)(x^2-2)) + (-22x^2-18x-8) < 0 . \qquad (4.2.37)$$

The first term in (4.2.37) is unchanged from (4.2.34). The roots of the second term are still imaginary, and hence the eigenvalue is negative. Therefore X is negative definite, and the solution to (4.2.22) is in fact a local maximum for D.

The locally D-optimal designs for $\beta_1=0(.02)1,2(1)5$ are presented in Table 4.2. The value of the original $D=\lambda_{11}\lambda_{22}$ is also given for comparison purposes. By comparing the present values of D to the corresponding values of D associated with the designs in Table 4.1, it can be seen that the present designs offer a slight improvement for $\beta_2=0$. Theorem 3.3 is applicable here since $\beta_2=0$, and so the optimal design for any particular value of β_1 is the same as the optimal design for $(-\beta_1)$. Recall that for a given β_1, the optimal design is to make half

of the total comparisons between x and 1, where x is found in the table, and the other half between (-x) and (-1). The optimal value of x is graphed as a function of $|\beta_1|$ for $|\beta_1| \leq 1$ in Figure 4.3.

Condition 3

A grid method was used to find D-optimal designs with no restrictions on the values of the two parameters. With $\beta_2 \neq 0$, the argument for a symmetric design is no longer applicable. As a generalization of the first subsection, the three comparisons $(-1,x_1)$, $(x_2,1)$, and $(-1,1)$ were allowed. A description of the grid method follows.

In general, if there are k independent unknowns, and the grid is to have n values for each variable at each stage, then the grid method is simply an evaluation of the function at each of the n^k points, and the determination of which of these points maximizes (or minimizes) the function. It then repeats the procedure, only with a finer grid centered at the best point previously found. This procedure is repeated until the optimal point has been found with sufficient accuracy.

The procedure used for the present condition was essentially a grid within a grid. At each step, the combination of three values each of x_1 and x_2 were taken. Initially these values were -0.5, 0, and 0.5. For each of these 9 points, a grid method was used on the allocation of the comparisons to the three pairs. This grid consisted of initial increments of 0.1 in the allocation fractions. Notice that this is not a full 11^3 grid since, for example, $\{\mu_{-1,x_1} = \mu_{x_2,1} = \mu_{-1,1} = 0.5\}$ is not a possible allocation for the three pairs because they must sum to 1. After the best of these possible allocations were found, a finer 5^2

grid was performed about the current best allocation. Note that it is a 5^2 grid rather than a 5^3 grid, because when two of the allocation proportions are known, the third is just the sum of the two proportions subtracted from 1. This was then repeated until the best allocation fractions were accurate to $\pm.0001$. The best of the 9 combinations of x_1 and x_2 was taken, and then a finer 3^2 grid about the current best combination of x_1 and x_2 was performed using the same procedure described in the present paragraph.

The entire procedure described in the preceding paragraph was repeated until the values of x_1 and x_2 were also accurate to $\pm.0001$. D-optimal designs were found for combinations of $\beta_1=0(.1)1,2,3$ and $\beta_2=0(.1)1,2,3$, and are presented in Table 4.3. Once again, the value of D for N=1 is also given in the table. Notice from the table that when $n_{x_2,1}=0$, the value of x_2 is immaterial, and therefore an asterisk appears under the column for x_2.

Theorem 4.1 is applicable, so designs for negative β_1 or β_2 can also be found from the table. If both β_1 and β_2 are negative, then the optimal design is identical to the one for positive β_1 and β_2. If, however, exactly one of the parameters is negative, then the optimal design is the same as the one for $|\beta_1|$ and $|\beta_2|$, except that n_{-1,x_1} changes to $n_{1,-x_1}$, and $n_{x_2,1}$ changes to $n_{-x_2,-1}$. For example, if $\beta_1=0.6$, $\beta_2=-0.3$, then the D-optimal design would be to run .462N comparisons of the pair (-1,.23), .468N of the pair (-.25,1), and .070N of the pair (-1,1).

Table 4.3 indicates that once again no pair should be compared more than half the time. Also, the optimal design is inside the region

Table 4.1. D-optimal designs with restriction x=-1,0,1

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	D	Efficiency
0	0	.333	.333	.333	.083333	.9613
.1	0	.334	.334	.332	.082506	.9613
.2	0	.337	.337	.327	.080098	.9613
.3	0	.341	.341	.318	.076314	.9612
.4	0	.348	.348	.305	.071460	.9608
.5	0	.357	.357	.287	.065895	.9599
.6	0	.369	.369	.263	.059986	.9583
.7	0	.384	.384	.232	.054068	.9555
.8	0	.405	.405	.190	.048426	.9536
.9	0	.433	.433	.135	.043283	.9542
1	0	.470	.470	.060	.038814	.9592
2	0	.500	.500	0	.011024	.9664
3	0	.500	.500	0	.002041	.7260
0	.1	.333	.333	.334	.083056	.9615
.1	.1	.336	.332	.332	.082233	.9617
.2	.1	.340	.333	.327	.079836	.9615
.3	.1	.345	.336	.319	.076070	.9613
.4	.1	.353	.342	.306	.071237	.9610
.5	.1	.362	.350	.288	.065694	.9602
.6	.1	.374	.362	.264	.059805	.9585
.7	.1	.389	.378	.233	.053905	.9558
.8	.1	.409	.399	.192	.048276	.9538
.9	.1	.435	.428	.137	.043141	.9543
1	.1	.470	.467	.063	.038676	.9591
2	.1	.500	.500	0	.011001	.9660
3	.1	.500	.500	0	.002039	.7257
0	.2	.332	.332	.336	.082233	.9619
.1	.2	.336	.330	.334	.081422	.9619
.2	.2	.342	.329	.329	.079060	.9619
.3	.2	.349	.331	.321	.075346	.9618
.4	.2	.357	.335	.308	.070575	.9615
.5	.2	.367	.343	.291	.065097	.9608
.6	.2	.378	.354	.268	.059269	.9594
.7	.2	.393	.370	.237	.053421	.9567
.8	.2	.411	.392	.197	.047833	.9544
.9	.2	.436	.421	.143	.042725	.9545
1	.2	.468	.461	.070	.038270	.9587
2	.2	.500	.500	0	.010931	.9647
3	.2	.500	.500	0	.002034	.7245
0	.3	.331	.331	.338	.080888	.9625
.1	.3	.337	.327	.337	.080098	.9625
.2	.3	.344	.324	.332	.077793	.9626

Table 4.1 - continued

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	D	Efficiency
.3	.3	.352	.324	.324	.074164	.9626
.4	.3	.361	.327	.312	.069495	.9625
.5	.3	.371	.334	.296	.064123	.9619
.6	.3	.382	.344	.274	.058396	.9607
.7	.3	.396	.360	.244	.052635	.9582
.8	.3	.413	.382	.205	.047114	.9556
.9	.3	.435	.412	.154	.042051	.9550
1	.3	.465	.452	.083	.037616	.9582
2	.3	.500	.500	0	.010817	.9624
3	.3	.500	.500	0	.002024	.7227
0	.4	.329	.329	.342	.079060	.9633
.1	.4	.337	.323	.341	.078297	.9634
.2	.4	.345	.319	.336	.076070	.9635
.3	.4	.354	.317	.329	.072558	.9637
.4	.4	.364	.318	.318	.068030	.9637
.5	.4	.374	.323	.303	.062805	.9634
.6	.4	.385	.333	.282	.057216	.9625
.7	.4	.398	.348	.254	.051575	.9604
.8	.4	.414	.369	.217	.046148	.9574
.9	.4	.433	.399	.168	.041149	.9559
1	.4	.460	.440	.101	.036744	.9578
2	.4	.500	.500	0	.010658	.9593
3	.4	.500	.500	0	.002011	.7201
0	.5	.327	.327	.346	.076799	.9643
.1	.5	.336	.319	.345	.076070	.9644
.2	.5	.347	.312	.342	.073940	.9647
.3	.5	.357	.308	.335	.070575	.9650
.4	.5	.367	.308	.325	.066222	.9653
.5	.5	.378	.311	.311	.061182	.9653
.6	.5	.389	.320	.292	.055768	.9647
.7	.5	.400	.334	.266	.050279	.9631
.8	.5	.414	.354	.232	.044972	.9599
.9	.5	.432	.383	.185	.040057	.9574
1	.5	.455	.423	.122	.035695	.9577
2	.5	.500	.500	0	.010456	.9551
3	.5	.500	.500	0	.001995	.7167
0	.6	.324	.324	.352	.074164	.9655
.1	.6	.336	.314	.351	.073475	.9656
.2	.6	.348	.305	.348	.071460	.9660
.3	.6	.359	.299	.342	.068267	.9665
.4	.6	.371	.296	.334	.064123	.9670
.5	.6	.381	.297	.321	.059302	.9674

Table 4.1 - continued

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	D	Efficiency
.6	.6	.392	.304	.304	.054097	.9672
.7	.6	.403	.317	.281	.048790	.9661
.8	.6	.415	.336	.249	.043628	.9634
.9	.6	.430	.364	.206	.038816	.9598
1	.6	.450	.403	.147	.034511	.9582
2	.6	.500	.500	0	.010213	.9500
3	.6	.500	.500	0	.001974	.7127
0	.7	.321	.321	.358	.071218	.9667
.1	.7	.335	.308	.357	.070575	.9669
.2	.7	.349	.297	.355	.068690	.9674
.3	.7	.362	.288	.350	.065694	.9681
.4	.7	.374	.283	.343	.061787	.9690
.5	.7	.385	.282	.333	.057216	.9697
.6	.7	.396	.286	.318	.052251	.9700
.7	.7	.406	.297	.297	.047154	.9695
.8	.7	.416	.315	.269	.042162	.9675
.9	.7	.429	.342	.229	.037472	.9634
1	.7	.445	.379	.176	.033239	.9598
2	.7	.500	.500	0	.009931	.9437
3	.7	.500	.500	0	.001950	.7079
0	.8	.318	.318	.364	.068030	.9680
.1	.8	.334	.302	.364	.067436	.9682
.2	.8	.350	.288	.362	.065694	.9689
.3	.8	.365	.276	.359	.062914	.9699
.4	.8	.378	.268	.354	.059269	.9710
.5	.8	.390	.264	.346	.054977	.9721
.6	.8	.400	.266	.334	.050279	.9729
.7	.8	.410	.275	.316	.045418	.9730
.8	.8	.419	.291	.291	.040618	.9718
.9	.8	.429	.315	.256	.036069	.9683
1	.8	.442	.351	.208	.031923	.9627
2	.8	.500	.500	0	.009613	.9363
3	.8	.500	.500	0	.001922	.7024
0	.9	.315	.315	.371	.064663	.9693
.1	.9	.334	.296	.371	.064123	.9696
.2	.9	.352	.278	.370	.062534	.9704
.3	.9	.369	.263	.369	.059986	.9716
.4	.9	.383	.251	.365	.056624	.9731
.5	.9	.396	.244	.360	.052635	.9746
.6	.9	.406	.243	.351	.048229	.9759
.7	.9	.415	.249	.336	.043628	.9765
.8	.9	.423	.262	.315	.039041	.9761

Table 4.1 - continued

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	D	Efficiency
.9	.9	.431	.285	.285	.034650	.9736
1	.9	.440	.318	.242	.030606	.9673
2	.9	.500	.500	0	.009262	.9277
3	.9	.500	.500	0	.001890	.6963
0	1	.311	.311	.378	.061182	.9705
.1	1	.333	.289	.378	.060698	.9709
.2	1	.354	.268	.378	.059269	.9719
.3	1	.373	.249	.379	.056966	.9734
.4	1	.389	.233	.378	.053905	.9752
.5	1	.403	.222	.375	.050239	.9772
.6	1	.414	.217	.369	.046148	.9789
.7	1	.422	.219	.359	.041828	.9800
.8	1	.429	.229	.342	.037472	.9802
.9	1	.435	.249	.316	.033256	.9786
1	1	.441	.279	.279	.029328	.9737
2	1	.500	.500	0	.008882	.9178
3	1	.500	.500	0	.001854	.6894
0	2	.279	.279	.441	.029328	.9631
.1	2	.357	.196	.448	.029392	.9649
.2	2	.430	.102	.468	.029580	.9704
.3	2	.500	0	.500	.029881	.9796
.4	2	.500	0	.500	.029897	.9863
.5	2	.500	0	.500	.029324	.9857
.6	2	.500	0	.500	.028229	.9833
.7	2	.500	0	.500	.026706	.9817
.8	2	.500	0	.500	.024863	.9809
.9	2	.500	0	.500	.022808	.9809
1	2	.500	0	.500	.020643	.9817
2	2	.491	.255	.255	.004495	.7529
3	2	.500	.500	0	.001307	.5856
0	3	.262	.262	.476	.011829	.8470
.1	3	.498	.002	.500	.012236	.8612
.2	3	.500	0	.500	.012983	.8864
.3	3	.500	0	.500	.013500	.9027
.4	3	.500	0	.500	.013767	.9099
.5	3	.500	0	.500	.013783	.9083
.6	3	.500	0	.500	.013565	.9024
.7	3	.500	0	.500	.013142	.8982
.8	3	.500	0	.500	.012551	.8955
.9	3	.500	0	.500	.011831	.8944
1	3	.500	0	.500	.011024	.8947
2	3	.500	0	.500	.003473	.7923
3	3	.499	.251	.251	.000618	.4284

Table 4.2.   D-optimal (local) designs with $\beta_2 = 0$

β_1	Optimal x	D	β_1	Optimal x	D
0	-.3333	.087792	.56	-.2742	.067635
.02	-.3333	.087760	.58	-.2701	.066450
.04	-.3330	.087667	.60	-.2660	.065255
.06	-.3326	.087511	.62	-.2617	.064054
.08	-.3321	.087294	.64	-.2572	.062849
.10	-.3314	.087016	.66	-.2527	.061641
.12	-.3305	.086678	.68	-.2481	.060432
.14	-.3295	.086282	.70	-.2434	.059224
.16	-.3283	.085827	.72	-.2386	.058019
.18	-.3270	.085317	.74	-.2337	.056819
.20	-.3255	.084751	.76	-.2287	.055624
.22	-.3238	.084133	.78	-.2237	.054437
.24	-.3221	.083464	.80	-.2185	.053258
.26	-.3201	.082745	.82	-.2133	.052090
.28	-.3180	.081979	.84	-.2080	.050932
.30	-.3158	.081168	.86	-.2026	.049787
.32	-.3134	.080315	.88	-.1972	.048655
.34	-.3109	.079421	.90	-.1917	.047537
.36	-.3082	.078490	.92	-.1861	.046434
.38	-.3054	.077523	.94	-.1805	.045346
.40	-.3025	.076523	.96	-.1748	.044275
.42	-.2994	.075492	.98	-.1691	.043221
.44	-.2962	.074434	1	-.1634	.042183
.46	-.2928	.073350	2	.1275	.011803
.48	-.2894	.072243	3	.3383	.003872
.50	-.2858	.071116	4	.4746	.001561
.52	-.2820	.069970	5	.5664	.000736
.54	-.2782	.068809			

Table 4.3. D-optimal designs

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	D
0	0	.24	-.24	.405	.405	.191	.090170
.1	0	.24	-.24	.406	.406	.188	.089275
.2	0	.24	-.24	.411	.411	.178	.086675
.3	0	.24	-.24	.419	.419	.162	.082605
.4	0	.24	-.24	.432	.432	.137	.077413
.5	0	.24	-.24	.450	.451	.099	.071512
.6	0	.25	-.25	.477	.477	.046	.065324
.7	0	.24	-.24	.500	.500	0	.059224
.8	0	.22	-.22	.500	.500	0	.053258
.9	0	.19	-.19	.500	.500	0	.047537
1	0	.16	-.16	.500	.500	0	.042183
2	0	-.13	.13	.500	.500	0	.011803
3	0	-.34	.34	.500	.500	0	.003872
0	.1	.24	-.24	.404	.404	.193	.089846
.1	.1	.24	-.24	.406	.405	.189	.088955
.2	.1	.24	-.24	.412	.409	.180	.086366
.3	.1	.24	-.24	.420	.417	.163	.082312
.4	.1	.24	-.24	.433	.430	.137	.077140
.5	.1	.24	-.24	.451	.448	.101	.071259
.6	.1	.25	-.24	.476	.475	.049	.065090
.7	.1	.25	-.24	.500	.500	0	.059009
.8	.1	.23	-.21	.500	.500	0	.053070
.9	.1	.20	-.18	.500	.500	0	.047375
1	.1	.17	-.15	.500	.500	0	.042046
2	.1	-.11	.15	.500	.500	0	.011789
3	.1	-.32	.35	.500	.500	0	.003872
0	.2	.24	-.24	.403	.403	.195	.088884
.1	.2	.24	-.24	.405	.403	.193	.088005
.2	.2	.24	-.24	.411	.405	.184	.085449
.3	.2	.24	-.24	.420	.413	.168	.081445
.4	.2	.24	-.23	.432	.425	.143	.076333
.5	.2	.24	-.24	.450	.443	.107	.070512
.6	.2	.25	-.24	.473	.470	.057	.064396
.7	.2	.26	-.24	.500	.500	0	.058369
.8	.2	.23	-.21	.500	.500	0	.052511
.9	.2	.21	-.18	.500	.500	0	.046895
1	.2	.18	-.14	.500	.500	0	.041638
2	.2	-.09	.17	.500	.500	0	.011747
3	.2	-.31	.37	.500	.500	0	.003874
0	.3	.24	-.24	.400	.400	.201	.087317
.1	.3	.24	-.24	.403	.399	.198	.086457
.2	.3	.24	-.24	.410	.401	.189	.083956

Table 4.3 - continued

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	D
.3	.3	.24	-.23	.418	.407	.175	.080035
.4	.3	.24	-.23	.430	.418	.152	.075019
.5	.3	.24	-.23	.446	.435	.119	.069298
.6	.3	.25	-.23	.468	.462	.070	.063272
.7	.3	.26	-.23	.500	.500	0	.057323
.8	.3	.24	-.20	.500	.500	0	.051597
.9	.3	.22	-.17	.500	.500	0	.046107
1	.3	.20	-.14	.500	.500	0	.040969
2	.3	-.07	.19	.500	.500	0	.011678
3	.3	-.29	.39	.500	.500	0	.003876
0	.4	.24	-.24	.396	.396	.208	.085195
.1	.4	.24	-.24	.401	.394	.205	.084362
.2	.4	.24	-.24	.407	.395	.198	.081937
.3	.4	.24	-.23	.416	.400	.184	.078128
.4	.4	.24	-.23	.427	.410	.163	.073246
.5	.4	.24	-.23	.442	.426	.132	.067663
.6	.4	.25	-.23	.462	.450	.088	.061763
.7	.4	.26	-.22	.490	.487	.024	.055914
.8	.4	.25	-.20	.500	.500	0	.050351
.9	.4	.23	-.16	.500	.500	0	.045032
1	.4	.21	-.13	.500	.500	0	.040056
2	.4	-.06	.21	.500	.500	0	.011582
3	.4	-.27	.40	.500	.500	0	.003879
0	.5	.24	-.24	.392	.392	.217	.082584
.1	.5	.24	-.24	.397	.389	.214	.081784
.2	.5	.23	-.24	.404	.388	.208	.079454
.3	.5	.23	-.23	.413	.392	.195	.075786
.4	.5	.23	-.23	.424	.399	.177	.071074
.5	.5	.24	-.23	.437	.413	.150	.065666
.6	.5	.24	-.22	.454	.435	.111	.059927
.7	.5	.26	-.21	.479	.469	.052	.054207
.8	.5	.26	-.20	.500	.500	0	.048806
.9	.5	.24	-.16	.500	.500	0	.043697
1	.5	.22	-.12	.500	.500	0	.038919
2	.5	-.04	.23	.500	.500	0	.011462
3	.5	-.26	.42	.500	.500	0	.003883
0	.6	.24	-.24	.387	.386	.227	.079560
.1	.6	.24	-.24	.393	.382	.225	.078800
.2	.6	.23	-.24	.401	.380	.219	.076581
.3	.6	.23	-.23	.409	.381	.210	.073082
.4	.6	.23	-.23	.420	.387	.194	.068571
.5	.6	.23	-.22	.432	.398	.170	.063372

Table 4.3 - continued

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	D
.6	.6	.24	-.21	.447	.417	.136	.057829
.7	.6	.25	-.20	.466	.448	.086	.052270
.8	.6	.27	-.19	.497	.496	.007	.047002
.9	.6	.26	-.15	.500	.500	0	.042133
1	.6	.24	-.11	.500	.500	0	.037585
2	.6	-.02	.25	.500	.500	0	.011317
3	.6	-.24	.44	.500	.500	0	.003887
0	.7	.24	-.24	.381	.381	.237	.076205
.1	.7	.24	-.24	.389	.375	.236	.075489
.2	.7	.23	-.24	.398	.371	.231	.073397
.3	.7	.23	-.23	.406	.370	.224	.070089
.4	.7	.23	-.23	.416	.373	.211	.065809
.5	.7	.23	-.22	.427	.380	.193	.060853
.6	.7	.23	-.21	.439	.396	.165	.055537
.7	.7	.24	-.19	.455	.422	.123	.050170
.8	.7	.26	-.18	.479	.465	.056	.045038
.9	.7	.27	-.14	.500	.500	0	.040377
1	.7	.25	-.09	.500	.500	0	.036083
2	.7	-.01	.27	.500	.500	0	.011152
3	.7	-.22	.45	.500	.500	0	.003891
0	.8	.24	-.24	.375	.375	.249	.072601
.1	.8	.24	-.24	.385	.368	.248	.071933
.2	.8	.23	-.24	.394	.361	.245	.069980
.3	.8	.23	-.24	.404	.357	.240	.066884
.4	.8	.22	-.23	.413	.356	.232	.062861
.5	.8	.22	-.22	.422	.361	.217	.058178
.6	.8	.22	-.20	.433	.371	.196	.053122
.7	.8	.23	-.18	.445	.393	.163	.047977
.8	.8	.24	-.16	.462	.428	.110	.043009
.9	.8	.28	-.13	.494	.489	.016	.038471
1	.8	.27	-.08	.500	.500	0	.034447
2	.8	.01	.29	.500	.500	0	.010966
3	.8	-.21	.47	.500	.500	0	.003895
0	.9	.24	-.24	.370	.370	.260	.068827
.1	.9	.24	-.25	.381	.359	.260	.068211
.2	.9	.23	-.24	.391	.350	.259	.066408
.3	.9	.22	-.24	.401	.343	.256	.063540
.4	.9	.22	-.23	.411	.338	.251	.059796
.5	.9	.21	-.21	.419	.338	.243	.055412
.6	.9	.21	-.19	.428	.343	.229	.050644
.7	.9	.21	-.17	.437	.359	.204	.045752
.8	.9	.22	-.14	.449	.387	.163	.040979

Table 4.3 - continued

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	D
.9	.9	.25	-.11	.470	.438	.092	.036555
1	.9	.29	-.06	.500	.500	0	.032711
2	.9	.03	.31	.500	.500	0	.010763
3	.9	-.19	.48	.500	.500	0	.003899
0	1	.25	-.25	.365	.365	.271	.064957
.1	1	.24	-.25	.377	.351	.272	.064396
.2	1	.23	-.25	.389	.339	.272	.062750
.3	1	.22	-.24	.400	.327	.273	.060123
.4	1	.21	-.23	.409	.319	.272	.056677
.5	1	.20	-.21	.418	.312	.270	.052615
.6	1	.20	-.19	.425	.312	.263	.048163
.7	1	.20	-.15	.432	.321	.247	.043553
.8	1	.20	-.11	.440	.342	.218	.039005
.9	1	.22	-.07	.453	.383	.164	.034727
1	1	.27	-.04	.482	.462	.056	.030935
2	1	.05	.34	.500	.500	0	.010544
3	1	-.18	.50	.500	.500	0	.003902
0	2	.46	-.46	.357	.357	.286	.031620
.1	2	.39	-.53	.396	.313	.292	.031568
.2	2	.32	-.60	.429	.260	.311	.031410
.3	2	.26	-.67	.458	.194	.349	.031136
.4	2	.21	-.74	.483	.104	.413	.030734
.5	2	.17	*	.500	0	.500	.030184
.6	2	.17	*	.500	0	.500	.029198
.7	2	.17	*	.500	0	.500	.027713
.8	2	.17	*	.500	0	.500	.025840
.9	2	.16	*	.500	0	.500	.023703
1	2	.15	*	.500	0	.500	.021419
2	2	.22	.53	.500	.500	0	.007930
3	2	-.02	.62	.500	.500	0	.003811
0	3	.71	-.71	.444	.444	.111	.016488
.1	3	.66	-.76	.459	.426	.115	.016497
.2	3	.61	-.81	.471	.399	.130	.016524
.3	3	.55	-.86	.482	.359	.159	.016568
.4	3	.50	-.91	.491	.279	.230	.016630
.5	3	.44	*	.500	0	.500	.016708
.6	3	.43	*	.500	0	.500	.016659
.7	3	.42	*	.500	0	.500	.016291
.8	3	.40	*	.500	0	.500	.015650
.9	3	.39	*	.500	0	.500	.014790
1	3	.38	*	.500	0	.500	.013771
2	3	.39	.66	.500	.500	0	.005532
3	3	.15	.71	.500	.500	0	.003369

depicted by Figure 4.2 for small parameter values, and the designs approach the boundary as the parameters become larger. Notice the peculiar designs for $\beta_2=2,3$ and $.5 \leq \beta_1 \leq 1$. The optimal design in this case is a comparison of only $(-1,x_1)$ and $(-1,1)$, each N/2 times.

A comparison of the designs in Table 4.2 with the designs for $\beta_2=0$ in Table 4.3 shows that at least some of the designs found in the previous subsection are only locally optimal. This follows from the fact that the value of D is larger in the present subsection than it is in the previous subsection for $\beta_1 \leq .6$. Notice that the designs are identical for $\beta_1 \geq .8$.

Based on what has been presented up to now, the design $D_1=\{n_{-1,.24} = n_{.24,1} = .4N, n_{-1,1} = .2N\}$ is a candidate for an overall efficient design. This is reasonably close to all of the optimal designs in Table 4.3 where the parameters are both less than or equal to 1. The efficiency of D_1 relative to the optimal designs is discussed in detail in Section 4.5. Also considered in Section 4.5 is the amount of bias present for design D_1 when the true model is cubic.

4.3 Average-variance Designs

The optimality criterion discussed in this section minimizes V, the average variance of the predicted value, where

$$V = \int_{-1}^{1} \text{var}(\ln \hat{\pi}_x) \, dx \quad . \tag{4.3.1}$$

The integral is over the experimental region, which in this case is $[-1,1]$, and $\ln \hat{\pi}_x = \hat{\beta}_1 x + \hat{\beta}_2 x^2$. The variance of the predicted value, $\ln \hat{\pi}_x$, is

$$\text{var}(\ln \hat{\pi}_x) = (\; x \quad x^2 \;) \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{pmatrix}^{-1} \begin{pmatrix} x \\ x^2 \end{pmatrix}$$

$$= \frac{x^2(\lambda_{22} - 2x\lambda_{12} + x^2\lambda_{11})}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} . \tag{4.3.2}$$

From (4.3.1) and (4.3.2),

$$V = \frac{\frac{2}{3}\lambda_{22} + \frac{2}{5}\lambda_{11}}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} . \tag{4.3.3}$$

Three methods for finding optimal designs analogous to the methods dis-
cussed in the previous section for D-optimal designs are now discussed.

Condition 1

The levels are again restricted to be $x=-1,0,1$. Because of the
complexity of V relative to D, it was decided to use a grid method for
this condition rather than approach the problem in a manner similar to
what was done for D-optimality. The grid consisted of initial incre-
ments of 0.1 in the allocation fractions. After finding the best of
the allocations above, finer and finer 5^2 grids were run until the
fractions were accurate to at least $\pm.001$.

The procedure described in the preceding paragraph was performed
for all combinations of $\beta_1=0(.1)1,2,3$ and $\beta_2=0(.1)1,2,3$. The designs
are presented in Table 4.4 along with the value of V for N=1. Again,
efficiencies of the "restricted" designs relative to the designs found
for Condition 3 are given. In this case, V is a multiple of $1/N$, so
the relative efficiency is defined to be the ratio of the corresponding

values of V. The relative efficiencies are larger than .95 when β_1 and β_2 are both less than or equal to 1, indicating that the designs in Table 4.4 are only slightly less efficient than the optimal designs found in Table 4.6.

The proof of Theorem 4.1 is again applicable, and so optimal designs for negative values of β_1 or β_2 can be found as before. That is, optimal designs for negative β_1 and β_2 are the same as for $|\beta_1|$ and $|\beta_2|$. When exactly one of the parameters is negative, the columns for $\mu_{-1,0}$ and $\mu_{0,1}$ effectively exchange.

Condition 2

In this subsection, local average-variance designs are found under the assumption that the true quadratic parameter is zero. The argument for symmetry is identical to the one used in Condition 2 for the D-optimality criterion. That is, reducing λ_{12}^2 reduces V, and so the optimal design must be symmetric because $\lambda_{12}=0$ for a symmetric design. Then from (4.3.3),

$$V \;=\; \frac{2/3}{\lambda_{11}} \;+\; \frac{2/5}{\lambda_{22}} \;. \tag{4.3.4}$$

Again, using a similar proof, it can be shown that every pair in the optimal design must include either x=-1 or x=1. Then from (4.3.4), V can be written

$$V \;=\; \frac{2/3}{f(\underline{x})} \;+\; \frac{2/5}{g(\underline{x})} \;, \tag{4.3.5}$$

where $f(\underline{x})$ and $g(\underline{x})$ are defined in (4.2.15) and (4.2.16).

The partial derivative of V with respect to x_i is

$$\frac{\partial V}{\partial x_i} = \frac{-\frac{2}{3}\frac{\partial f(\underline{x})}{\partial x_i}}{f^2(\underline{x})} + \frac{-\frac{2}{5}\frac{\partial g(\underline{x})}{\partial x_i}}{g^2(\underline{x})} \quad , \tag{4.3.6}$$

where $\partial f(\underline{x})/\partial x_i$ and $\partial g(\underline{x})/\partial x_i$ are given in (4.2.19) and (4.2.20). After evaluating (4.3.6) at the point $x_i = x$, for all i, setting it equal to zero, and multiplying by $N^2(1-x^2)^3/8\cosh$, the locally optimal design is found by solving

$$(\frac{2}{3}(1+x)^3 + \frac{4}{5}x)\cosh + \left[\frac{(1-x)(1+x)^3}{3} + \frac{(1-x^2)}{5}\right]\beta_1 \sinh = 0 . \tag{4.3.7}$$

The average-variance design is then to use half of the total available comparisons to compare 1 with x, where x is the solution to (4.3.7), and the other half to compare (-1) with $(-x)$. Again, a computer program is necessary to solve (4.3.7) for several values of β_1.

It needs to be shown that the matrix X is positive definite, where

$$X = \left(\frac{\partial^2 V}{\partial x_i \partial x_j}\bigg|_{x_i = x} \right) .$$

From (4.3.6), the second order partial derivatives are

$$\frac{\partial^2 V}{\partial x_i^2} = -\frac{2}{3}\left[\frac{\left[\frac{\partial^2 f(\underline{x})}{\partial x_i^2}\right]}{f^2(\underline{x})} - \frac{2\left[\frac{\partial f(\underline{x})}{\partial x_i}\right]^2}{f^3(\underline{x})}\right] - \frac{2}{5}\left[\frac{\left[\frac{\partial^2 g(\underline{x})}{\partial x_i^2}\right]}{g^2(\underline{x})} - \frac{2\left[\frac{\partial g(\underline{x})}{\partial x_i}\right]^2}{g^3(\underline{x})}\right] ,$$

$$\tag{4.3.8}$$

$$i = 1, \ldots, N/2,$$

$$\frac{\partial^2 v}{\partial x_i \partial x_j} \;=\; \frac{\dfrac{4}{3}\,\dfrac{\partial f(\underline{x})}{\partial x_i}\,\dfrac{\partial f(\underline{x})}{\partial x_j}}{f^3(\underline{x})} \;+\; \frac{\dfrac{4}{5}\,\dfrac{\partial g(\underline{x})}{\partial x_i}\,\dfrac{\partial g(\underline{x})}{\partial x_j}}{g^3(\underline{x})} \;,\quad \begin{array}{l} i \neq j, \\ i,j=1,\ldots,N/2, \end{array}$$

$$(4.3.9)$$

where (4.3.9) follows from the fact that

$$\frac{\partial^2 f(\underline{x})}{\partial x_i \partial x_j} \;=\; \frac{\partial^2 g(\underline{x})}{\partial x_i \partial x_j} \;=\; 0\;. \qquad\qquad (4.3.10)$$

Using the same notation of the previous section, $(\frac{N}{2} - 1)$ of the eigenvalues of x are

$$a - b \;=\; \left[\frac{-\dfrac{2}{3}\,\dfrac{\partial^2 f(\underline{x})}{\partial x_i^{\,2}}}{f^2(\underline{x})} \;-\; \frac{\dfrac{2}{5}\,\dfrac{\partial^2 g(\underline{x})}{\partial x_i^{\,2}}}{g^2(\underline{x})}\right]\Bigg|_{x_i = x} \qquad (4.3.11)$$

By factoring out $4/N^2$,

$$a - b \;>\; 0$$

if and only if

$$\left\{ \begin{array}{l} -\dfrac{2}{3}\,\dfrac{\left[\begin{array}{c}(2 - \frac{1}{2}(1-x)^2\beta_1^{\,2})\cosh^2 + 4(1-x)\beta_1\cosh\cdot\sinh \\ -\,\frac{3}{2}(1-x)^2\beta_1^{\,2}\sinh^2\end{array}\right]}{(1-x)^4} \\[3em] -\dfrac{2}{5}\,\dfrac{\left[\begin{array}{c}(4(3x^2-1) - \frac{1}{2}(1-x^2)^2\beta_1^{\,2})\cosh^2 \\ +\,8(x-x^3)\beta_1\cosh\cdot\sinh + \frac{3}{2}(1-x^2)^2\beta_1^{\,2}\sinh^2\end{array}\right]}{(1-x^2)^4} \end{array} \right\} \;>\; 0\;.$$

$$(4.3.12)$$

The remaining eigenvalue is

$$(a-b) + \frac{N}{2}b = (a-b) + \frac{N}{2}\left[\frac{\left[\frac{4}{3}\left.\frac{\partial f(\underline{x})}{\partial x_i}\right|_{x_i=x}\right]^2}{f(\underline{x})} + \frac{\left[\frac{4}{5}\left.\frac{\partial g(\underline{x})}{\partial x_i}\right|_{x_i=x}\right]^2}{g(\underline{x})}\right] \cdot \quad (4.3.13)$$

Once again factoring out $4/N^2$,

$$(a - b) + \frac{N}{2}b > 0$$

if and only if

$$\left[\begin{array}{l} \frac{N^2}{4}(a-b) + \frac{4}{3}(-2(1-x)\cosh - (1-x)^2\beta_1\sinh)^2/(1-x)^6 \\ \\ + \frac{4}{5}(4(x^3-x)\cosh - (1-x^2)^2\beta_1\sinh)^2/(1-x^2)^6 \end{array}\right] > 0, \quad (4.3.14)$$

where $\frac{N^2}{4}(a-b)$ is given by (4.3.12).

An attempt was made to use the relationship between x and β_1 found in (4.3.7) to simplify (4.3.12) and (4.3.14) in a manner similar to what was done for D-optimal designs. However, it did not appear that a useful substitution could be extracted from (4.3.7). An APL computer program solved (4.3.7) for $\beta_1=0(.02)1,2(1)5$, and simultaneously evaluated both eigenvalues. They were both positive in all cases, and so the designs presented are at least locally optimal. The designs are presented in Table 4.5 along with the value of V. The optimal value of x is graphed as a function of $|\beta_1|$ in Figure 4.3. Again, Theorem 3.3 is applicable, and so the optimal designs for $(-\beta_1)$ and β_1 are identical.

Table 4.4. Average-variance designs with restriction x=-1,0,1

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	V	Efficiency
0	0	.382	.382	.236	3.655522	.9562
.1	0	.382	.382	.235	3.671145	.9563
.2	0	.383	.383	.235	3.718160	.9565
.3	0	.384	.383	.233	3.796988	.9568
.4	0	.385	.385	.231	3.908278	.9571
.5	0	.387	.387	.226	4.052855	.9570
.6	0	.390	.390	.220	4.231559	.9563
.7	0	.395	.395	.210	4.445071	.9549
.8	0	.402	.402	.197	4.693577	.9535
.9	0	.411	.411	.178	4.976284	.9525
1	0	.425	.425	.150	5.290571	.9524
2	0	.500	.500	0	10.159344	.9918
3	0	.500	.500	0	23.610992	.8026
0	.1	.382	.382	.236	3.662509	.9563
.1	.1	.382	.382	.236	3.678149	.9564
.2	.1	.382	.383	.235	3.725227	.9567
.3	.1	.382	.385	.233	3.804166	.9570
.4	.1	.383	.386	.231	3.915628	.9572
.5	.1	.385	.389	.226	4.060450	.9572
.6	.1	.387	.393	.220	4.239491	.9566
.7	.1	.391	.399	.211	4.453456	.9552
.8	.1	.396	.407	.197	4.702573	.9538
.9	.1	.405	.417	.178	4.986072	.9527
1	.1	.417	.433	.151	5.301409	.9526
2	.1	.481	.519	0	10.184770	.9909
3	.1	.477	.523	0	23.670074	.8012
0	.2	.382	.382	.236	3.683503	.9567
.1	.2	.382	.382	.236	3.699206	.9568
.2	.2	.382	.383	.235	3.746469	.9571
.3	.2	.381	.385	.234	3.825734	.9575
.4	.2	.382	.387	.231	3.937707	.9578
.5	.2	.382	.391	.227	4.083259	.9579
.6	.2	.384	.395	.221	4.263307	.9574
.7	.2	.386	.402	.212	4.478637	.9562
.8	.2	.390	.411	.199	4.729557	.9547
.9	.2	.397	.423	.181	5.015437	.9535
1	.2	.407	.439	.154	5.333891	.9532
2	.2	.462	.538	0	10.261279	.9880
3	.2	.455	.545	0	23.847900	.7969
0	.3	.382	.382	.237	3.718629	.9572
.1	.3	.381	.383	.237	3.734430	.9573
.2	.3	.381	.384	.236	3.781997	.9577

Table 4.4 - continued

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	V	Efficiency
.3	.3	.380	.386	.235	3.861807	.9582
.4	.3	.379	.389	.232	3.974609	.9587
.5	.3	.379	.392	.229	4.121360	.9590
.6	.3	.380	.397	.223	4.303068	.9588
.7	.3	.381	.405	.214	4.520638	.9577
.8	.3	.384	.414	.202	4.774541	.9562
.9	.3	.389	.426	.185	5.064346	.9549
1	.3	.397	.444	.159	5.387944	.9541
2	.3	.443	.557	0	10.389654	.9831
3	.3	.433	.567	0	24.146240	.7899
0	.4	.382	.382	.237	3.768089	.9580
.1	.4	.381	.382	.237	3.784023	.9581
.2	.4	.380	.383	.237	3.832005	.9586
.3	.4	.378	.386	.236	3.912556	.9592
.4	.4	.378	.389	.234	4.026499	.9600
.5	.4	.377	.393	.231	4.174885	.9605
.6	.4	.377	.398	.226	4.358873	.9607
.7	.4	.376	.406	.218	4.579528	.9599
.8	.4	.378	.416	.206	4.837540	.9584
.9	.4	.381	.429	.190	5.132753	.9568
1	.4	.387	.448	.165	5.463449	.9556
2	.4	.425	.575	0	10.571164	.9764
3	.4	.411	.589	0	24.568100	.7803
0	.5	.381	.381	.238	3.832173	.9589
.1	.5	.380	.382	.238	3.848271	.9591
.2	.5	.379	.383	.238	3.896767	.9596
.3	.5	.378	.386	.237	3.978236	.9605
.4	.5	.376	.389	.236	4.093597	.9615
.5	.5	.374	.393	.233	4.244030	.9624
.6	.5	.373	.398	.229	4.430865	.9630
.7	.5	.372	.407	.222	4.655393	.9627
.8	.5	.372	.417	.211	4.918571	.9613
.9	.5	.373	.431	.196	5.220604	.9595
1	.5	.377	.449	.174	5.560246	.9576
2	.5	.407	.593	0	10.807646	.9676
3	.5	.389	.611	0	25.117706	.7681
0	.6	.381	.381	.239	3.911265	.9599
.1	.6	.379	.382	.239	3.927553	.9601
.2	.6	.378	.383	.239	3.976645	.9609
.3	.6	.377	.385	.239	4.059182	.9619
.4	.6	.375	.388	.237	4.176199	.9633
.5	.6	.372	.392	.236	4.329037	.9646

Table 4.4 - continued

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	V	Efficiency
.6	.6	.370	.398	.232	4.519240	.9656
.7	.6	.367	.406	.227	4.748350	.9660
.8	.6	.365	.417	.218	5.017673	.9650
.9	.6	.365	.431	.204	5.327829	.9629
1	.6	.367	.449	.184	5.678129	.9604
2	.6	.389	.611	0	11.101451	.9570
3	.6	.368	.632	0	25.800507	.7535
0	.7	.380	.380	.240	4.005838	.9610
.1	.7	.379	.382	.240	4.022338	.9613
.2	.7	.378	.383	.240	4.072088	.9622
.3	.7	.376	.384	.240	4.155814	.9635
.4	.7	.373	.387	.240	4.274682	.9652
.5	.7	.370	.391	.238	4.430226	.9670
.6	.7	.367	.397	.236	4.624232	.9685
.7	.7	.363	.405	.232	4.858556	.9695
.8	.7	.360	.415	.225	5.134883	.9693
.9	.7	.358	.430	.213	5.454336	.9672
1	.7	.357	.448	.195	5.816876	.9639
2	.7	.372	.628	0	11.455528	.9444
3	.7	.348	.652	0	26.623413	.7369
0	.8	.380	.380	.240	4.116484	.9621
.1	.8	.379	.381	.240	4.133205	.9625
.2	.8	.378	.382	.240	4.183656	.9636
.3	.8	.375	.383	.242	4.268641	.9652
.4	.8	.372	.386	.242	4.389500	.9673
.5	.8	.369	.389	.242	4.547972	.9695
.6	.8	.365	.395	.241	4.746127	.9716
.7	.8	.360	.403	.237	4.986179	.9732
.8	.8	.356	.413	.232	5.270257	.9738
.9	.8	.351	.427	.222	5.600039	.9724
1	.8	.349	.444	.207	5.976202	.9685
2	.8	.355	.645	0	11.873416	.9299
3	.8	.328	.672	0	27.594589	.7182
0	.9	.380	.380	.240	4.243881	.9633
.1	.9	.379	.380	.241	4.260830	.9637
.2	.9	.377	.381	.242	4.311986	.9649
.3	.9	.375	.382	.243	4.398276	.9669
.4	.9	.372	.384	.244	4.521192	.9694
.5	.9	.368	.386	.246	4.682725	.9721
.6	.9	.363	.392	.245	4.885263	.9748
.7	.9	.357	.399	.244	5.131424	.9770
.8	.9	.352	.409	.240	5.423844	.9782

Table 4.4 - continued

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	V	Efficiency
.9	.9	.346	.422	.232	5.764808	.9777
1	.9	.342	.440	.219	6.155793	.9740
2	.9	.340	.661	0	12.359278	.9135
3	.9	.309	.691	0	28.723801	.6979
0	1	.380	.380	.240	4.388850	.9643
.1	1	.379	.380	.241	4.406016	.9648
.2	1	.378	.379	.242	4.457863	.9663
.3	1	.376	.380	.244	4.545424	.9686
.4	1	.372	.381	.247	4.670382	.9715
.5	1	.368	.383	.249	4.835003	.9748
.6	1	.363	.387	.251	5.042024	.9780
.7	1	.356	.394	.250	5.294521	.9807
.8	1	.349	.403	.248	5.595672	.9825
.9	1	.342	.415	.243	5.948477	.9826
1	1	.335	.432	.233	6.355247	.9800
2	1	.324	.676	0	12.918030	.8954
3	1	.291	.709	0	30.022354	.6762
0	2	.388	.387	.225	7.091832	.9494
.1	2	.400	.372	.228	7.106284	.9499
.2	2	.410	.354	.235	7.150141	.9515
.3	2	.420	.333	.248	7.224962	.9541
.4	2	.428	.306	.266	7.333524	.9580
.5	2	.436	.275	.290	7.480139	.9631
.6	2	.442	.239	.319	7.671077	.9693
.7	2	.447	.198	.355	7.915125	.9761
.8	2	.449	.155	.396	8.224180	.9827
.9	2	.447	.116	.437	8.613729	.9891
1	2	.439	.083	.478	9.103029	.9857
2	2	.194	.563	.243	23.599945	.6616
3	2	.156	.844	0	56.220154	.4295
0	3	.410	.408	.182	13.868884	.7998
.1	3	.461	.351	.188	13.832186	.8029
.2	3	.519	.274	.207	13.711627	.8128
.3	3	.602	.149	.249	13.468477	.8324
.4	3	.691	0	.309	13.028968	.8676
.5	3	.672	0	.328	12.626303	.9848
.6	3	.651	0	.348	12.353458	.9940
.7	3	.629	0	.371	12.208490	.9940
.8	3	.603	0	.397	12.193017	.9943
.9	3	.578	0	.422	12.311801	.9952
1	3	.551	0	.449	12.572087	.9966
2	3	.268	0	.732	28.217896	.7686
3	3	.086	.666	.248	129.685013	.2324

Table 4.5. Average-variance (local) designs with $\beta_2=0$

β_1	Optimal x	V	β_1	Optimal x	V
0	-.2937	3.509664	.56	-.2570	3.976627
.02	-.2936	3.510251	.58	-.2544	4.011062
.04	-.2935	3.512013	.60	-.2517	4.046773
.06	-.2932	3.514951	.62	-.2490	4.083764
.08	-.2929	3.519064	.64	-.2461	4.122042
.10	-.2925	3.524355	.66	-.2432	4.161614
.12	-.2919	3.530823	.68	-.2403	4.202487
.14	-.2913	3.538471	.70	-.2372	4.244667
.16	-.2906	3.547300	.72	-.2341	4.288160
.18	-.2898	3.557312	.74	-.2309	4.332974
.20	-.2889	3.568509	.76	-.2276	4.379116
.22	-.2879	3.580894	.78	-.2243	4.426592
.24	-.2868	3.594469	.80	-.2208	4.475409
.26	-.2856	3.609237	.82	-.2174	4.525575
.28	-.2843	3.625202	.84	-.2138	4.577096
.30	-.2829	3.642367	.86	-.2102	4.629979
.32	-.2815	3.660736	.88	-.2065	4.684231
.34	-.2799	3.680311	.90	-.2028	4.739859
.36	-.2782	3.701098	.92	-.1990	4.796871
.38	-.2765	3.723100	.94	-.1951	4.855272
.40	-.2747	3.746323	.96	-.1912	4.915070
.42	-.2728	3.770770	.98	-.1872	4.976272
.44	-.2708	3.796447	1	-.1832	5.038884
.46	-.2687	3.823359	2	.0599	10.076305
.48	-.2665	3.851511	3	.2832	18.950738
.50	-.2642	3.880908	4	.4370	31.519167
.52	-.2619	3.911555	5	.5402	47.666037
.54	-.2595	3.943460			

Table 4.6. Average-variance designs

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	V
0	0	.24	-.24	.456	.456	.088	3.495387
.1	0	.24	-.24	.457	.456	.087	3.510654
.2	0	.24	-.24	.459	.459	.082	3.556512
.3	0	.24	-.24	.463	.463	.074	3.633069
.4	0	.24	-.24	.470	.470	.061	3.740429
.5	0	.24	-.24	.480	.480	.040	3.878488
.6	0	.25	-.25	.495	.495	.010	4.046638
.7	0	.24	-.24	.500	.500	0	4.244664
.8	0	.22	-.22	.500	.500	0	4.475407
.9	0	.20	-.20	.500	.500	0	4.739857
1	0	.18	-.18	.500	.500	0	5.038882
2	0	-.06	.06	.500	.500	0	10.076301
3	0	-.28	.28	.500	.500	0	18.950729
0	.1	.24	-.24	.455	.455	.089	3.502490
.1	.1	.24	-.24	.455	.457	.088	3.517797
.2	.1	.24	-.24	.456	.461	.083	3.563775
.3	.1	.24	-.24	.459	.466	.075	3.640539
.4	.1	.24	-.24	.464	.474	.062	3.748207
.5	.1	.24	-.24	.473	.486	.041	3.886702
.6	.1	.25	-.24	.486	.503	.012	4.055433
.7	.1	.24	-.23	.490	.510	0	4.254045
.8	.1	.23	-.21	.489	.511	0	4.485379
.9	.1	.21	-.19	.488	.512	0	4.750462
1	.1	.19	-.17	.487	.513	0	5.050146
2	.1	-.03	.09	.486	.515	0	10.091684
3	.1	-.26	.31	.486	.514	0	18.964417
0	.2	.24	-.24	.455	.455	.090	3.523829
.1	.2	.24	-.24	.453	.458	.089	3.539256
.2	.2	.24	-.24	.453	.463	.085	3.585588
.3	.2	.24	-.24	.454	.468	.077	3.662973
.4	.2	.24	-.24	.458	.478	.064	3.771568
.5	.2	.24	-.24	.464	.490	.046	3.911353
.6	.2	.25	-.24	.475	.508	.017	4.081825
.7	.2	.25	-.23	.481	.519	0	4.282252
.8	.2	.24	-.21	.478	.522	0	4.515371
.9	.2	.22	-.19	.477	.523	0	4.782364
1	.2	.21	-.16	.475	.525	0	5.084027
2	.2	-.01	.11	.471	.529	0	10.137798
3	.2	-.24	.33	.472	.528	0	19.005569
0	.3	.24	-.24	.453	.453	.094	3.559501
.1	.3	.24	-.24	.451	.457	.092	3.575117
.2	.3	.24	-.24	.449	.463	.088	3.622039

Table 4.6 - continued

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	V
.3	.3	.24	-.23	.449	.470	.081	3.700450
.4	.3	.24	-.23	.451	.480	.070	3.810571
.5	.3	.24	-.23	.455	.493	.052	3.952484
.6	.3	.25	-.23	.463	.511	.026	4.125826
.7	.3	.25	-.23	.471	.529	0	4.329484
.8	.3	.24	-.20	.468	.532	0	4.565605
.9	.3	.23	-.18	.465	.535	0	4.835806
1	.3	.22	-.15	.463	.537	0	5.140771
2	.3	.01	.14	.456	.544	0	10.214447
3	.3	-.21	.35	.458	.542	0	19.074036
0	.4	.24	-.24	.452	.452	.097	3.609651
.1	.4	.24	-.24	.448	.457	.096	3.625533
.2	.4	.24	-.24	.445	.463	.092	3.673269
.3	.4	.24	-.23	.444	.471	.086	3.753095
.4	.4	.24	-.23	.443	.481	.076	3.865322
.5	.4	.24	-.23	.445	.494	.061	4.010170
.6	.4	.25	-.23	.450	.513	.037	4.187459
.7	.4	.26	-.22	.461	.538	.001	4.396066
.8	.4	.25	-.20	.457	.543	0	4.636456
.9	.4	.24	-.17	.454	.546	0	4.911190
1	.4	.23	-.14	.451	.550	0	5.220799
2	.4	.04	.16	.442	.558	0	10.321324
3	.4	-.19	.37	.444	.556	0	19.169724
0	.5	.24	-.24	.450	.450	.101	3.674506
.1	.5	.24	-.24	.445	.456	.099	3.690723
.2	.5	.23	-.24	.441	.462	.097	3.739486
.3	.5	.23	-.23	.438	.471	.092	3.821097
.4	.5	.23	-.23	.436	.481	.084	3.935978
.5	.5	.24	-.22	.435	.494	.072	4.084516
.6	.5	.24	-.22	.437	.512	.051	4.266764
.7	.5	.26	-.21	.444	.536	.020	4.481932
.8	.5	.26	-.20	.447	.553	0	4.728442
.9	.5	.25	-.16	.442	.558	0	5.009092
1	.5	.24	-.13	.438	.562	0	5.324689
2	.5	.06	.19	.428	.572	0	10.457992
3	.5	-.17	.39	.431	.569	0	19.292358
0	.6	.24	-.24	.447	.448	.105	3.754354
.1	.6	.24	-.24	.442	.454	.104	3.770972
.2	.6	.23	-.24	.437	.462	.102	3.820965
.3	.6	.23	-.23	.432	.469	.099	3.904708
.4	.6	.23	-.23	.428	.479	.093	4.022751
.5	.6	.23	-.22	.425	.493	.083	4.175665

Table 4.6 - continued

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	V
.6	.6	.24	-.21	.424	.509	.067	4.363788
.7	.6	.25	-.20	.427	.532	.041	4.586706
.8	.6	.27	-.19	.437	.563	0	4.842248
.9	.6	.26	-.16	.431	.569	0	5.130244
1	.6	.26	-.12	.427	.574	0	5.453180
2	.6	.08	.22	.414	.586	0	10.623909
3	.6	-.14	.41	.417	.583	0	19.441772
0	.7	.24	-.24	.446	.446	.109	3.849566
.1	.7	.24	-.24	.439	.452	.109	3.866647
.2	.7	.23	-.24	.432	.459	.109	3.918055
.3	.7	.23	-.23	.426	.468	.106	4.004243
.4	.7	.23	-.23	.420	.477	.102	4.125902
.5	.7	.23	-.22	.415	.490	.095	4.283813
.6	.7	.23	-.21	.411	.505	.084	4.478602
.7	.7	.24	-.19	.410	.525	.066	4.710282
.8	.7	.26	-.18	.415	.554	.032	4.977325
.9	.7	.28	-.15	.420	.580	0	5.275561
1	.7	.27	-.10	.415	.585	0	5.607162
2	.7	.11	.24	.401	.599	0	10.818448
3	.7	-.12	.43	.405	.595	0	19.617645
0	.8	.24	-.24	.443	.443	.114	3.960589
.1	.8	.24	-.24	.436	.450	.114	3.978189
.2	.8	.23	-.24	.429	.458	.114	4.031176
.3	.8	.23	-.24	.421	.465	.114	4.120092
.4	.8	.22	-.23	.414	.475	.112	4.245771
.5	.8	.22	-.22	.406	.485	.109	4.409204
.6	.8	.22	-.20	.399	.498	.103	4.611306
.7	.8	.23	-.18	.394	.515	.091	4.852514
.8	.8	.24	-.16	.393	.540	.067	5.131998
.9	.8	.28	-.13	.402	.579	.018	5.445721
1	.8	.29	-.09	.403	.597	0	5.787655
2	.8	.13	.27	.388	.612	0	11.040878
3	.8	-.10	.45	.392	.608	0	19.819672
0	.9	.24	-.24	.441	.441	.118	4.087960
.1	.9	.24	-.25	.434	.449	.118	4.106125
.2	.9	.23	-.24	.425	.456	.119	4.160837
.3	.9	.22	-.24	.417	.463	.121	4.252724
.4	.9	.22	-.23	.408	.471	.122	4.382762
.5	.9	.21	-.21	.398	.479	.123	4.552135
.6	.9	.21	-.19	.389	.489	.122	4.762028
.7	.9	.21	-.17	.380	.503	.117	5.013263
.8	.9	.22	-.14	.375	.523	.102	5.305647

113

Table 4.6 - continued

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	V
.9	.9	.25	-.10	.375	.555	.069	5.636459
1	.9	.31	-.07	.392	.608	0	5.995749
2	.9	.15	.30	.376	.624	0	11.290431
3	.9	-.08	.47	.379	.621	0	20.047470
0	1	.25	-.25	.439	.440	.121	4.232290
.1	1	.24	-.25	.431	.447	.122	4.251060
.2	1	.23	-.25	.423	.453	.124	4.307615
.3	1	.22	-.24	.413	.460	.127	4.402675
.4	1	.21	-.23	.403	.466	.132	4.537338
.5	1	.20	-.21	.392	.472	.136	4.712965
.6	1	.20	-.19	.380	.480	.141	4.930936
.7	1	.20	-.15	.369	.489	.142	5.192368
.8	1	.20	-.11	.359	.505	.136	5.497597
.9	1	.22	-.07	.354	.529	.117	5.845107
1	1	.27	-.04	.360	.577	.063	6.228132
2	1	.17	.32	.364	.636	0	11.566305
3	1	-.06	.48	.368	.633	0	20.300705
0	2	.46	-.46	.490	.490	.021	6.733291
.1	2	.39	-.53	.493	.489	.018	6.750583
.2	2	.32	-.60	.497	.492	.010	6.803201
.3	2	.26	-.66	.499	.501	0	6.893327
.4	2	.20	-.72	.496	.504	0	7.025329
.5	2	.14	-.76	.493	.507	0	7.204350
.6	2	.10	-.81	.488	.512	0	7.435870
.7	2	.05	-.85	.482	.518	0	7.725970
.8	2	.01	-.88	.476	.524	0	8.081570
.9	2	-.02	.34	.416	.176	.409	8.520247
1	2	-.03	.37	.399	.173	.428	8.972814
2	2	.35	.53	.269	.731	0	15.613154
3	2	.14	.62	.270	.730	0	24.145706
0	3	.60	-.60	.500	.500	0	11.092550
.1	3	.55	-.64	.504	.497	0	11.105566
.2	3	.50	-.69	.507	.493	0	11.144773
.3	3	.45	-.73	.511	.489	0	11.210546
.4	3	.40	-.78	.516	.484	0	11.303552
.5	3	.24	.53	.471	.249	.281	12.434281
.6	3	.16	.54	.581	.070	.349	12.279836
.7	3	.12	*	.608	0	.392	12.135128
.8	3	.11	*	.584	0	.416	12.123116
.9	3	.09	*	.560	0	.440	12.253007
1	3	.08	*	.536	0	.464	12.528843
2	3	.48	.66	.209	.791	0	21.688461
3	3	.29	.71	.208	.792	0	30.141708

Condition 3

A grid method identical to the one described in Condition 3 for D-optimal designs is used to minimize V in (4.3.3). The optimal designs are presented in Table 4.6. Theorem 4.1 is again applicable, and so optimal designs for negative values of β_1 or β_2 can be found in the manner described in Condition 3 of Section 4.2.

Notice that the pattern of the designs in Table 4.6 is similar to the D-optimal designs in Table 4.3. However, it is no longer true that every pair is compared no more than half the time. Also, for $\beta_2 \leq 1$, the change in x_1, x_2, μ_{-1,x_1}, $\mu_{x_2,1}$, or $\mu_{-1,1}$ found in Table 4.6 is monotone as a function of β_1. For $\beta_2=2$ or $\beta_2=3$, however, this is not the case. For example, notice the large difference between the optimal design for $\beta_1=.8$, $\beta_2=2$ and the optimal design for $\beta_1=.9$, $\beta_2=2$.

Once again a comparison of the designs in Table 4.5 with the designs for $\beta_2=0$ in Table 4.6 shows the designs found in the preceding subsection result in only local minimums for V when $|\beta_1| \leq .6$. The corresponding designs found presently are superior because of the smaller value of V.

4.4 Minimax Designs

This section is a consideration of designs which minimize the maximum value of $\mathrm{var}(\ln \hat{\pi}_x)$ over $x \in [-1,1]$. The designs are referred to as minimax designs. From (4.3.2), the minimax design minimizes

$$\max_{x \in [-1,1]} \mathrm{var}(\ln \hat{\pi}_x) = \max_{x \in [-1,1]} \left[\frac{x^2(\lambda_{22} - 2x\lambda_{12} + x^2\lambda_{11})}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \right] . \quad (4.4.1)$$

The following theorem simplifies (4.4.1).

Theorem 4.2. Given a quadratic model. Then $\mathrm{var}(\ln \hat{\pi}_x)$ is a maximum over $x\varepsilon[-1,1]$ at $x=\pm 1$.

Proof. Note that

$$
\mathrm{var}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \;=\; \frac{\begin{pmatrix} \lambda_{22} & -\lambda_{12} \\ -\lambda_{12} & \lambda_{11} \end{pmatrix}}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \quad .
$$

Case 1: $\lambda_{12} \le 0$.

$$
\begin{aligned}
\mathrm{var}(\ln \hat{\pi}_1) \;&=\; \mathrm{var}(\hat{\beta}_1 + \hat{\beta}_2) \\[2mm]
&=\; \frac{\lambda_{22} + \lambda_{11} - 2\lambda_{12}}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \\[2mm]
&\ge\; \frac{x^2(\lambda_{22} + \lambda_{11}x^2 - 2\lambda_{12}x)}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \qquad (-1 \le x \le 1) \\[2mm]
&=\; \mathrm{var}(\ln \hat{\pi}_x)
\end{aligned}
$$

Case 2: $\lambda_{12} > 0$.

$$
\begin{aligned}
\mathrm{var}(\ln \hat{\pi}_{-1}) \;&=\; \mathrm{var}(\hat{\beta}_2 - \hat{\beta}_1) \\[2mm]
&=\; \frac{\lambda_{11} + \lambda_{22} + 2\lambda_{12}}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \\[2mm]
&\ge\; \frac{x^2(\lambda_{11}x^2 + \lambda_{22} + 2\lambda_{12}x)}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \qquad (-1 \le x \le 1)
\end{aligned}
$$

$$= \quad \text{var}(\ln \hat{\pi}_x) \quad . \qquad \dagger$$

From the preceding theorem, the minimax design is the design which minimizes

$$M \quad = \quad \max \left[\frac{\lambda_{22} - 2\lambda_{12} + \lambda_{11}}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \quad , \quad \frac{\lambda_{22} + 2\lambda_{12} + \lambda_{11}}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \right]$$

$$= \quad \frac{\lambda_{22} + 2|\lambda_{12}| + \lambda_{11}}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \quad . \qquad (4.4.2)$$

Three conditions are again briefly discussed. The procedures for the three conditions are very similar to the procedures described for average-variance designs because of the similarity of V and M given by (4.3.3) and (4.4.2), respectively.

Condition 1

The levels are restricted to be x=-1,0,1. A grid method identical to the one described in Condition 1 of Section 4.3 is used. The proof of Theorem 4.1 is applicable, and so optimal designs for negative β_1 or β_2 are found as explained for Condition 1 of Section 4.3. These "restricted" minimax designs are presented in Table 4.7 along with the values of M, for N=1. The efficiencies of the "restricted" designs relative to the designs for Condition 3 are once again given. In this case, M is a multiple of 1/N, so the relative efficiency is defined to be the ratio of the corresponding values of M. The relative efficiencies are once again larger than 0.95 when β_1 and β_2 are both less than or equal to 1, indicating that these designs are only slightly less

efficient than the optimal designs found at the end of the present section in Table 4.9.

Condition 2

In this subsection local minimax designs are found under the assumption that the true quadratic parameter is zero. The same argument for symmetry as used in Condition 2 of Section 4.2 can be applied here. Decreasing $\lambda_{12}{}^2$ reduces the numerator of (4.4.2) and increases the denominator, and hence decreases M. So the minimax design must be a symmetric design, implying $\lambda_{12}=0$. Then from (4.4.2),

$$M \quad = \quad \frac{1}{\lambda_{11}} \quad + \quad \frac{1}{\lambda_{22}} \quad . \qquad (4.4.3)$$

Also, the same argument used in Condition 2 of Section 4.2 can again be applied to show that every pair in the optimal design must be of the form $(-1,x_i)$ or $(x_i,1)$, some x_i, and hence M can be written

$$M \quad = \quad \frac{1}{f(\underline{x})} \quad + \quad \frac{1}{g(\underline{x})} \quad , \qquad (4.4.4)$$

where $f(\underline{x})$ and $g(\underline{x})$ are defined in (4.2.15) and (4.2.16). Notice that the only differences between (4.4.4) and (4.3.5) are the coefficients 2/3 and 2/5.

The partial derivative of M with respect to x_i is

$$\frac{\partial M}{\partial x_i} \quad = \quad \frac{-\dfrac{\partial f(\underline{x})}{\partial x_i}}{f^2(\underline{x})} \quad + \quad \frac{-\dfrac{\partial g(\underline{x})}{\partial x_i}}{g^2(\underline{x})} \quad , \qquad (4.4.5)$$

where $\partial f(\underline{x})/\partial x_i$ and $\partial g(\underline{x})/\partial x_i$ are given in (4.2.19) and (4.2.20).

Via the same procedure employed in Condition 2 of Section 4.3, the optimal design is a comparison of $(-1,-x)$ and $(x,1)$ an equal number of times, where x is the solution to

$$(2(1+x)^3+4x)\cosh + ((1-x)(1+x)^3+(1-x^2))\beta_1\sinh = 0 \qquad (4.4.6)$$

To show that the matrix of second order partial derivatives is positive definite, equations $(4.3.8)-(4.3.14)$ can be repeated for the present situation. Everything is identical except that a multiple of 2/3 and 2/5 is removed from the places where they appear. The eigen-value (a-b) can then be shown to be equal to (4.3.11) after the coefficients 2/3 and 2/5 are removed from (4.3.11). Likewise, the eigen-value $(a-b) + \frac{N}{2}b$ is equal to (4.3.13) after the coefficients 2/3 and 2/5 are removed from (4.3.13).

An APL computer program solved (4.4.6) for $\beta_1=0(.02)1,2(1)5$, and simultaneously evaluated both eigenvalues. Both eigenvalues were posi-tive in all cases, and so the designs presented are at least locally minimax. The designs are presented in Table 4.8 along with the value of M. Again, Theorem 3.3 is applicable, and so the optimal design for a negative value of β_1 is the same as the optimal design for $|\beta_1|$.

Condition 3

A grid method identical to the one described in Condition 3 of Section 4.2 was used. The minimax designs are presented in Table 4.9. Theorem 4.1 is again applicable, and so optimal designs for negative values of β_1 or β_2 can be found in the manner described for Condition 3 of Section 4.2.

Table 4.7.   Minimax designs with restriction x=-1,0,1

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	M	Efficiency
0	0	.423	.423	.155	7.464108	.9517
.1	0	.423	.423	.154	7.491435	.9518
.2	0	.424	.424	.152	7.573516	.9521
.3	0	.426	.426	.149	7.710603	.9526
.4	0	.428	.428	.144	7.902992	.9533
.5	0	.431	.431	.137	8.150852	.9543
.6	0	.437	.437	.127	8.453897	.9555
.7	0	.443	.443	.114	8.810935	.9572
.8	0	.453	.453	.094	9.219111	.9596
.9	0	.467	.467	.067	9.672845	.9633
1	0	.485	.485	.030	10.161998	.9691
2	0	.500	.500	0	19.048782	.9723
3	0	.500	.500	0	44.270630	.7547
0	.1	.423	.423	.155	7.479891	.9518
.1	.1	.420	.424	.156	7.507998	.9518
.2	.1	.420	.429	.151	7.591027	.9521
.3	.1	.418	.431	.152	7.728451	.9526
.4	.1	.421	.437	.142	7.921003	.9535
.5	.1	.420	.441	.139	8.170156	.9544
.6	.1	.427	.452	.121	8.475105	.9556
.7	.1	.426	.456	.118	8.833076	.9573
.8	.1	.432	.466	.103	9.244843	.9596
.9	.1	.450	.489	.061	9.700654	.9633
1	.1	.459	.504	.037	10.192365	.9691
2	.1	.462	.538	0	19.128830	.9712
3	.1	.455	.545	0	44.485809	.7527
0	.2	.422	.422	.156	7.527333	.9521
.1	.2	.418	.426	.156	7.556258	.9521
.2	.2	.416	.433	.151	7.638617	.9526
.3	.2	.413	.438	.149	7.778711	.9531
.4	.2	.407	.441	.152	7.974294	.9540
.5	.2	.409	.451	.140	8.225651	.9551
.6	.2	.411	.461	.128	8.534287	.9563
.7	.2	.413	.472	.115	8.898054	.9581
.8	.2	.408	.475	.118	9.321956	.9597
.9	.2	.424	.502	.074	9.779487	.9639
1	.2	.439	.528	.033	10.281797	.9693
2	.2	.425	.575	0	19.356262	.9682
3	.2	.411	.589	0	45.092941	.7473
0	.3	.422	.422	.156	7.606749	.9528
.1	.3	.415	.427	.158	7.636298	.9528
.2	.3	.410	.435	.154	7.720152	.9533

Table 4.7 - continued

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	M	Efficiency
.3	.3	.403	.440	.157	7.863355	.9539
.4	.3	.396	.446	.158	8.064491	.9547
.5	.3	.401	.464	.135	8.319720	.9561
.6	.3	.397	.472	.131	8.634922	.9575
.7	.3	.400	.489	.110	9.008231	.9591
.8	.3	.393	.493	.113	9.437747	.9612
.9	.3	.389	.501	.110	9.930961	.9629
1	.3	.417	.549	.035	10.432550	.9695
2	.3	.388	.612	0	19.731598	.9637
3	.3	.368	.632	0	46.113266	.7385
0	.4	.421	.421	.157	7.718635	.9536
.1	.4	.412	.428	.161	7.748528	.9537
.2	.4	.407	.440	.153	7.835239	.9543
.3	.4	.395	.444	.161	7.983413	.9549
.4	.4	.395	.462	.143	8.186843	.9562
.5	.4	.382	.463	.155	8.454483	.9572
.6	.4	.386	.486	.127	8.777301	.9589
.7	.4	.370	.482	.148	9.174763	.9594
.8	.4	.380	.514	.106	9.603919	.9629
.9	.4	.386	.540	.074	10.100954	.9658
1	.4	.378	.545	.077	10.652193	.9692
2	.4	.353	.647	0	20.271484	.9570
3	.4	.327	.673	0	47.553284	.7265
0	.5	.421	.421	.159	7.863725	.9546
.1	.5	.412	.433	.156	7.894499	.9548
.2	.5	.397	.438	.165	7.985312	.9554
.3	.5	.387	.448	.165	8.138353	.9561
.4	.5	.382	.463	.155	8.348237	.9578
.5	.5	.372	.474	.154	8.624695	.9593
.6	.5	.370	.492	.139	8.960258	.9611
.7	.5	.369	.514	.117	9.361121	.9628
.8	.5	.363	.527	.110	9.822090	.9650
.9	.5	.365	.553	.082	10.342136	.9678
1	.5	.367	.578	.056	10.913309	.9710
2	.5	.320	.680	0	20.969727	.9485
3	.5	.289	.711	0	49.423553	.7118
0	.6	.420	.420	.161	8.042955	.9558
.1	.6	.410	.435	.156	8.075470	.9560
.2	.6	.393	.442	.165	8.170047	.9567
.3	.6	.379	.451	.169	8.329222	.9577
.4	.6	.372	.468	.160	8.547760	.9597
.5	.6	.369	.491	.140	8.834993	.9615

Table 4.7 - continued

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	M	Efficiency
.6	.6	.350	.492	.158	9.192108	.9630
.7	.6	.349	.517	.135	9.605767	.9661
.8	.6	.351	.547	.103	10.093208	.9673
.9	.6	.339	.555	.106	10.641331	.9696
1	.6	.342	.589	.069	11.246311	.9727
2	.6	.289	.712	0	21.845016	.9380
3	.6	.254	.747	0	51.742035	.6947
0	.7	.419	.419	.162	8.257525	.9571
.1	.7	.407	.436	.157	8.291222	.9573
.2	.7	.387	.443	.170	8.390544	.9582
.3	.7	.378	.462	.160	8.553052	.9601
.4	.7	.366	.477	.157	8.785346	.9622
.5	.7	.351	.489	.161	9.088071	.9640
.6	.7	.344	.509	.147	9.456905	.9671
.7	.7	.337	.532	.131	9.898378	.9687
.8	.7	.331	.554	.115	10.411327	.9712
.9	.7	.319	.565	.117	10.997437	.9722
1	.7	.312	.585	.103	11.652453	.9735
2	.7	.260	.740	0	22.891083	.9259
3	.7	.221	.779	0	54.549164	.6754
0	.8	.418	.418	.164	8.508865	.9584
.1	.8	.405	.437	.158	8.545020	.9586
.2	.8	.385	.448	.168	8.647012	.9600
.3	.8	.369	.463	.169	8.821605	.9617
.4	.8	.359	.485	.156	9.063863	.9644
.5	.8	.351	.510	.138	9.386874	.9667
.6	.8	.335	.523	.142	9.773139	.9700
.7	.8	.317	.529	.154	10.245109	.9724
.8	.8	.315	.564	.121	10.785250	.9746
.9	.8	.307	.586	.107	11.406818	.9761
1	.8	.300	.609	.091	12.101690	.9770
2	.8	.233	.767	0	24.133331	.9118
3	.8	.193	.807	0	57.868576	.6543
0	.9	.417	.417	.165	8.798699	.9597
.1	.9	.398	.433	.169	8.835464	.9601
.2	.9	.384	.454	.162	8.944130	.9616
.3	.9	.363	.466	.171	9.128388	.9638
.4	.9	.350	.489	.162	9.385105	.9669
.5	.9	.338	.512	.151	9.721140	.9702
.6	.9	.315	.517	.168	10.142849	.9730
.7	.9	.305	.540	.155	10.635491	.9766
.8	.9	.298	.569	.133	11.213647	.9794

Table 4.7 - continued

β_1	β_2	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,1}$	M	Efficiency
.9	.9	.288	.591	.122	11.879495	.9804
1	.9	.295	.649	.056	12.669651	.9771
2	.9	.209	.791	0	25.575104	.8960
3	.9	.167	.833	0	61.731598	.6320
0	1	.417	.417	.167	9.129013	.9609
.1	1	.396	.434	.171	9.167790	.9615
.2	1	.380	.457	.163	9.283387	.9632
.3	1	.362	.477	.162	9.476922	.9658
.4	1	.338	.488	.174	9.753860	.9690
.5	1	.328	.517	.155	10.107888	.9732
.6	1	.317	.545	.138	10.556965	.9766
.7	1	.294	.550	.156	11.080395	.9810
.8	1	.286	.582	.131	11.701517	.9840
.9	1	.272	.598	.130	12.414202	.9855
1	1	.263	.625	.112	13.222181	.9845
2	1	.187	.813	0	27.223267	.8787
3	1	.144	.856	0	66.143311	.6089
0	2	.416	.416	.168	15.363714	.9469
.1	2	.383	.446	.171	15.441003	.9468
.2	2	.354	.480	.165	15.671208	.9466
.3	2	.321	.506	.173	16.059067	.9471
.4	2	.293	.538	.169	16.609390	.9488
.5	2	.265	.564	.171	17.330353	.9525
.6	2	.240	.591	.170	18.233276	.9587
.7	2	.214	.609	.177	19.333313	.9675
.8	2	.191	.630	.179	20.642090	.9720
.9	2	.174	.660	.166	22.173538	.9561
1	2	.154	.672	.174	23.953873	.9378
2	2	.066	.933	0	60.633728	.6313
3	2	.033	.967	0	155.540634	.3733
0	3	.430	.430	.140	31.329956	.8058
.1	3	.392	.470	.139	31.497498	.8051
.2	3	.352	.506	.142	31.999695	.8033
.3	3	.316	.544	.140	32.844849	.8004
.4	3	.280	.577	.143	34.046616	.7967
.5	3	.251	.600	.149	35.632156	.7920
.6	3	.605	0	.395	35.907074	.8238
.7	3	.581	0	.419	35.822800	.8704
.8	3	.555	0	.445	36.144714	.9118
.9	3	.528	0	.472	36.891647	.9447
1	3	.500	0	.500	38.097580	.9682
2	3	.231	0	.769	95.645172	.6100
3	3	.010	.990	0	409.475830	.2012

Table 4.8. Minimax (local) designs with $\beta_2=0$

β_1	Optimal x	M	β_1	Optimal x	M
0	-.2291	7.103231	.56	-.1920	7.951115
.02	-.2290	7.104304	.58	-.1894	8.013255
.04	-.2289	7.107523	.60	-.1867	8.077642
.06	-.2286	7.112889	.62	-.1840	8.144283
.08	-.2283	7.120402	.64	-.1812	8.213184
.10	-.2279	7.130064	.66	-.1783	8.284353
.12	-.2273	7.141875	.68	-.1753	8.357794
.14	-.2267	7.155837	.70	-.1723	8.433517
.16	-.2260	7.171951	.72	-.1692	8.511525
.18	-.2251	7.190220	.74	-.1660	8.591828
.20	-.2242	7.210647	.76	-.1628	8.674431
.22	-.2232	7.233232	.78	-.1595	8.759341
.24	-.2221	7.257980	.80	-.1561	8.846566
.26	-.2209	7.284892	.82	-.1527	8.936111
.28	-.2196	7.313973	.84	-.1492	9.027984
.30	-.2182	7.345226	.86	-.1457	9.122192
.32	-.2167	7.378654	.88	-.1421	9.218741
.34	-.2151	7.414262	.90	-.1384	9.317639
.36	-.2134	7.452053	.92	-.1347	9.418892
.38	-.2117	7.492032	.94	-.1309	9.522507
.40	-.2098	7.534204	.96	-.1271	9.628489
.42	-.2079	7.578573	.98	-.1233	9.736847
.44	-.2059	7.625144	1	-.1193	9.847587
.46	-.2038	7.673922	2	.1066	18.521632
.48	-.2016	7.724913	3	.3087	33.412753
.50	-.1993	7.778122	4	.4510	54.321313
.52	-.1970	7.833555	5	.5487	81.078453
.54	-.1945	7.891217			

Table 4.9. Minimax designs

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	M
0	0	.23	-.23	.500	.500	0	7.103233
.1	0	.23	-.23	.500	.500	0	7.130062
.2	0	.22	-.22	.500	.500	0	7.210647
.3	0	.22	-.22	.500	.500	0	7.345225
.4	0	.21	-.21	.500	.500	0	7.534202
.5	0	.20	-.20	.500	.500	0	7.778128
.6	0	.19	-.19	.500	.500	0	8.077641
.7	0	.17	-.17	.500	.500	0	8.433517
.8	0	.16	-.16	.500	.500	0	8.846562
.9	0	.14	-.14	.500	.500	0	9.317636
1	0	.12	-.12	.500	.500	0	9.847586
2	0	-.11	.11	.500	.500	0	18.521622
3	0	-.31	.31	.500	.500	0	33.412704
0	.1	.23	-.23	.500	.500	0	7.119185
.1	.1	.23	-.23	.497	.503	0	7.146172
.2	.1	.22	-.23	.495	.506	0	7.227188
.3	.1	.22	-.22	.492	.509	0	7.362472
.4	.1	.21	-.21	.489	.511	0	7.552416
.5	.1	.20	-.20	.486	.514	0	7.797619
.6	.1	.19	-.19	.483	.517	0	8.098640
.7	.1	.17	-.17	.481	.519	0	8.456331
.8	.1	.16	-.15	.478	.522	0	8.871431
.9	.1	.15	-.13	.476	.524	0	9.344682
1	.1	.13	-.11	.474	.526	0	9.877072
2	.1	-.09	.12	.466	.534	0	18.576996
3	.1	-.29	.33	.467	.533	0	33.484009
0	.2	.23	-.23	.500	.500	0	7.167143
.1	.2	.23	-.23	.494	.506	0	7.194577
.2	.2	.22	-.23	.489	.511	0	7.276844
.3	.2	.22	-.22	.483	.517	0	7.414238
.4	.2	.21	-.21	.477	.523	0	7.607140
.5	.2	.20	-.20	.472	.528	0	7.856112
.6	.2	.19	-.18	.467	.533	0	8.161740
.7	.2	.18	-.17	.462	.538	0	8.524799
.8	.2	.16	-.15	.457	.543	0	8.945994
.9	.2	.15	-.13	.453	.547	0	9.426017
1	.2	.14	-.10	.449	.551	0	9.965639
2	.2	-.07	.15	.432	.568	0	18.741409
3	.2	-.27	.35	.434	.566	0	33.697693
0	.3	.23	-.23	.500	.500	0	7.247409
.1	.3	.23	-.23	.492	.508	0	7.275542
.2	.3	.23	-.23	.483	.517	0	7.359927

Table 4.9 - continued

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	M
.3	.3	.22	-.22	.474	.526	0	7.500859
.4	.3	.22	-.21	.465	.535	0	7.699010
.5	.3	.21	-.19	.457	.543	0	7.954224
.6	.3	.20	-.18	.450	.550	0	8.267556
.7	.3	.18	-.17	.443	.557	0	8.639672
.8	.3	.17	-.15	.436	.564	0	9.071136
.9	.3	.16	-.12	.430	.570	0	9.562373
1	.3	.14	-.10	.424	.576	0	10.114105
2	.3	-.04	.17	.399	.601	0	19.015671
3	.3	-.24	.37	.403	.598	0	34.052536
0	.4	.24	-.24	.500	.500	0	7.360502
.1	.4	.23	-.24	.489	.511	0	7.389644
.2	.4	.23	-.23	.477	.523	0	7.477092
.3	.4	.22	-.23	.467	.533	0	7.623173
.4	.4	.21	-.22	.456	.544	0	7.828116
.5	.4	.21	-.20	.445	.555	0	8.092491
.6	.4	.20	-.18	.434	.566	0	8.416834
.7	.4	.19	-.16	.424	.576	0	8.802041
.8	.4	.18	-.14	.415	.585	0	9.247813
.9	.4	.16	-.12	.407	.593	0	9.755398
1	.4	.15	-.09	.400	.600	0	10.323983
2	.4	-.02	.19	.367	.633	0	19.398895
3	.4	-.22	.39	.372	.628	0	34.546707
0	.5	.24	-.24	.498	.498	.004	7.507002
.1	.5	.24	-.24	.484	.513	.003	7.537487
.2	.5	.23	-.24	.472	.527	.001	7.628940
.3	.5	.23	-.23	.457	.543	0	7.781449
.4	.5	.22	-.22	.444	.556	0	7.995684
.5	.5	.20	-.21	.432	.567	.002	8.273508
.6	.5	.21	-.18	.417	.581	.001	8.611589
.7	.5	.20	-.16	.406	.594	0	9.012951
.8	.5	.19	-.13	.394	.606	0	9.478048
.9	.5	.19	-.09	.383	.617	0	10.008749
1	.5	.16	-.08	.376	.624	0	10.597091
2	.5	0	.21	.337	.663	0	19.890823
3	.5	-.20	.40	.342	.658	0	35.178162
0	.6	.24	-.24	.495	.495	.011	7.687564
.1	.6	.23	-.23	.475	.509	.016	7.719963
.2	.6	.23	-.24	.463	.530	.007	7.816233
.3	.6	.23	-.23	.449	.549	.002	7.977216
.4	.6	.23	-.22	.433	.567	0	8.203011
.5	.6	.22	-.20	.417	.583	0	8.494607

Table 4.9 - continued

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	M
.6	.6	.20	-.20	.404	.596	0	8.851893
.7	.6	.22	-.18	.388	.612	0	9.279871
.8	.6	.19	-.14	.376	.624	0	9.763409
.9	.6	.18	-.11	.364	.636	0	10.318142
1	.6	.19	-.09	.353	.647	0	10.939160
2	.6	.03	.23	.309	.691	0	20.490387
3	.6	-.18	.42	.314	.686	0	35.944672
0	.7	.24	-.24	.491	.491	.018	7.903124
.1	.7	.24	-.23	.473	.513	.014	7.937562
.2	.7	.24	-.24	.457	.534	.010	8.039942
.3	.7	.22	-.24	.439	.550	.011	8.211526
.4	.7	.25	-.22	.420	.578	.002	8.452837
.5	.7	.22	-.21	.405	.595	0	8.760774
.6	.7	.19	-.19	.383	.600	.017	9.145382
.7	.7	.21	-.17	.372	.628	0	9.588939
.8	.7	.18	-.15	.359	.641	.001	10.111272
.9	.7	.20	-.10	.342	.658	0	10.692100
1	.7	.19	-.06	.331	.669	0	11.343455
2	.7	.06	.26	.283	.717	0	21.195862
3	.7	-.15	.44	.288	.712	0	36.841553
0	.8	.24	-.24	.488	.488	.024	8.154837
.1	.8	.23	-.24	.468	.508	.024	8.191410
.2	.8	.23	-.24	.447	.532	.021	8.300882
.3	.8	.22	-.24	.427	.553	.020	8.484043
.4	.8	.22	-.23	.408	.576	.016	8.741317
.5	.8	.20	-.23	.391	.596	.013	9.074366
.6	.8	.21	-.19	.371	.621	.008	9.479525
.7	.8	.20	-.16	.350	.633	.017	9.962827
.8	.8	.21	-.13	.338	.662	0	10.511288
.9	.8	.20	-.10	.324	.677	0	11.134350
1	.8	.21	-.05	.309	.691	0	11.823655
2	.8	.08	.28	.259	.741	0	22.005203
3	.8	-.13	.45	.264	.736	0	37.865631
0	.9	.24	-.24	.486	.486	.029	8.444048
.1	.9	.24	-.24	.461	.508	.031	8.483164
.2	.9	.23	-.24	.439	.532	.029	8.600701
.3	.9	.23	-.23	.413	.554	.033	8.797866
.4	.9	.23	-.23	.395	.582	.024	9.074232
.5	.9	.21	-.21	.373	.602	.025	9.431239
.6	.9	.20	-.19	.353	.624	.023	9.869430
.7	.9	.21	-.17	.337	.653	.011	10.386945
.8	.9	.20	-.12	.315	.667	.018	10.982839

Table 4.9 - continued

β_1	β_2	x_1	x_2	μ_{-1,x_1}	$\mu_{x_2,1}$	$\mu_{-1,1}$	M
.9	.9	.22	-.09	.304	.696	0	11.646238
1	.9	.22	-.04	.289	.711	0	12.379728
2	.9	.12	.30	.238	.762	0	22.915192
3	.9	-.10	.47	.242	.758	0	39.012115
0	1	.25	-.25	.483	.483	.034	8.772314
.1	1	.24	-.25	.459	.507	.034	8.814472
.2	1	.24	-.24	.429	.533	.037	8.941481
.3	1	.22	-.24	.407	.555	.038	9.152744
.4	1	.21	-.24	.385	.582	.034	9.451067
.5	1	.20	-.21	.359	.603	.038	9.836727
.6	1	.20	-.19	.336	.627	.037	10.310200
.7	1	.19	-.15	.314	.648	.038	10.870255
.8	1	.21	-.12	.298	.682	.020	11.514256
.9	1	.21	-.07	.279	.696	.025	12.233892
1	1	.24	-.03	.270	.730	.001	13.016977
2	1	.15	.33	.218	.782	0	23.921951
3	1	-.07	.49	.222	.778	0	40.275284
0	2	.41	-.41	.500	.500	0	14.548218
.1	2	.34	-.48	.480	.520	0	14.618847
.2	2	.27	-.53	.458	.542	0	14.834950
.3	2	.21	-.56	.433	.567	0	15.209146
.4	2	.14	-.60	.407	.593	0	15.759338
.5	2	.08	-.62	.381	.619	0	16.507782
.6	2	.03	-.63	.352	.648	0	17.479965
.7	2	-.03	-.64	.326	.674	0	18.704865
.8	2	0	.31	.174	.624	.202	20.063736
.9	2	0	.34	.156	.637	.207	21.200836
1	2	-.02	.37	.144	.642	.213	22.464966
2	2	.47	.53	.126	.874	0	38.277344
3	2	.24	.62	.117	.883	0	58.069946
0	3	.57	-.57	.500	.500	0	25.244232
.1	3	.52	-.61	.477	.523	0	25.358597
.2	3	.46	-.65	.455	.545	0	25.704498
.3	3	.40	-.68	.434	.566	0	26.289764
.4	3	.32	-.71	.417	.583	0	27.124908
.5	3	.22	-.74	.405	.595	0	28.220917
.6	3	.10	-.76	.403	.597	0	29.581024
.7	3	-.10	-.79	.421	.579	0	31.180283
.8	3	-.13	-.83	.434	.566	0	32.955276
.9	3	-.14	-.87	.446	.555	0	34.851456
1	3	-.16	-.91	.456	.544	0	36.885345
2	3	.65	.66	.100	.900	0	58.347656
3	3	.47	.71	.089	.911	0	82.387863

The optimal values of x_1 and x_2 in Table 4.9 are once again simi-
lar to the previous two optimality criteria discussed in the present
chapter. However, notice that the pair (-1,1) should seldom or never
be compared for the minimax criterion. That is the major difference
between the present designs and the designs in the two previous sections.
The optimal designs for $\beta_2=0$ found in Table 4.9 are identical to the
designs found in Table 4.8, suggesting that the minimax designs found
in Table 4.8 may in fact give absolute minimums for M.

4.5 Conclusion and Design Recommendations

Figure 4.3 is a graph of the optimal level to be compared with
x=1 for each of the three criteria discussed in the present chapter
when $\beta_2=0$, as a function of $|\beta_1|$. This figure has been referred to in
Condition 2 for each of the three previous sections. Recall that these
are only locally optimal in some cases, and hence are possibly not as
useful as the optimal designs found for Condition 3 in the three pre-
vious sections. However, notice that the optimal value of x is closer
to 0 for the minimax criterion than it is for the other two criteria.
Also, the slope of the D-optimal curve is steeper than the other two
which are essentially parallel, signifying that these D-optimal designs
change faster than the others as $|\beta_1|$ changes.

The remainder of this section contains a discussion and comparison
of seven particular designs on how good they are overall. Estimates
of their relative efficiencies are obtained for the three optimality
criteria discussed in the present chapter, and it is determined which
designs are best overall for protecting against possible bias.

Figure 4.3. Locally optimal designs when $\beta_2=0$

Let Designs 1-3 be defined as three point, four point, and five point designs, respectively, where the levels are equally spaced, and all pairs are compared an equal number of times. Also, define Designs 4-7 as:

Design 4: $\{\mu_{-1,.24} = \mu_{-.24,1} = .4,\ \mu_{-1,1} = .2\}$,

Design 5: $\{\mu_{-1,.24} = \mu_{-.24,1} = .45,\ \mu_{-1,1} = .1\}$,

Design 6: $\{\mu_{-1,.2} = \mu_{-.2,1} = .5\}$,

Design 7: $\{\mu_{-1,.5} = \mu_{-.5,1} = .4706,\ \mu_{-1,1} = .0588\}$.

The reason for including Designs 1-3 is that they are designs which might naively be run by an experimenter who has little knowledge of optimal designs. The motivation for including Designs 4 and 5 is that they are similar to many of the optimal designs found in Tables 4.3, 4.6, and 4.9, especially for relatively small parameter values. Design 6 is similar to the optimal designs for large parameter values. The source of Design 7 is Springall's paper. It is included here although he did not state any optimal properties for the design.

The efficiencies of the seven designs defined above relative to the optimal designs found in Tables 4.3, 4.6, and 4.9 were calculated for combinations of $\beta_1, \beta_2 = 0(.2)1$. The integrated mean square error (J) was calculated for $\beta_1, \beta_2 = 0(.25)1$ and $\beta_3 = -1(.4)1$, where

$$J = \int_{-1}^{1} E(\ln \hat{\pi}_x - g(x))^2\ dx\ ,$$

N=100, and the true model, g(x), is cubic. Finally, the bias part of J was calculated. See the following chapter for an expression for J and the bias.

Table 4.10 presents the minimum, maximum, and median of the relative efficiencies for each of the three criteria discussed in the present chapter. This table gives an overall estimate of what is lost by running a design different from the optimal design. The estimates are given for the 24 combinations of β_1 and β_2 where $|\beta_1| \leq .6$, for the 12 where $|\beta_1| > .6$, and for all 36 combined. Remember that these minimums, maximums, and medians are only estimates of what could be considered as corresponding true parameter values.

The efficiencies for the D-optimal criterion will now be discussed. The average-variance and minimax criteria can be considered in a similar manner. For the condition $|\beta_1| \leq 1$, Designs 4 and 5 both have a median over 99%, with Design 4 slightly larger. However Design 5 has a minimum efficiency of 96.6% as opposed to 93.5% for Design 4. Both of these designs are overall superior to the other designs considered. For the case $|\beta_1| \leq .6$, Designs 4 and 5 are again superior to the others. However, Design 6 is the better design for the case $|\beta_1| > .6$, although Design 5 is very close.

The values of the bias and the integrated mean square error were ranked from smallest (1) to largest (7) for each of the 150 combinations of β_1, β_2, and β_3 considered. Table 4.11 presents an arithmetic mean of these ranks. This gives a rough idea of how the bias and integrated mean square error compare among the seven designs.

It is no surprise that Design 3 has a small bias since there are a total of ten pairs compared an equal number of times. Surprisingly, Design 7 had the smallest bias every time. However, the variance part of J is much larger for Designs 3 and 7 than for Designs 5 and 6, and

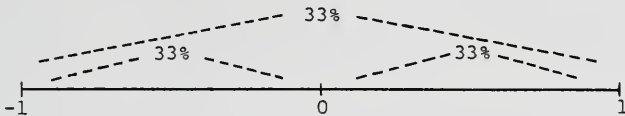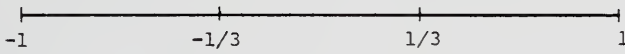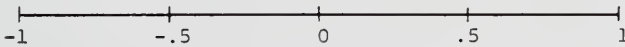hence the latter two designs have smaller values of J as indicated by the average rankings in Table 4.11.
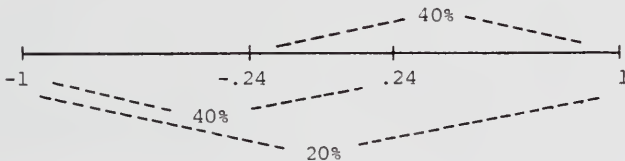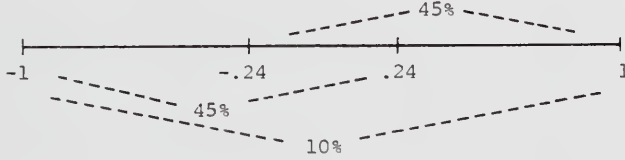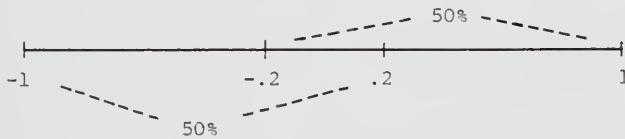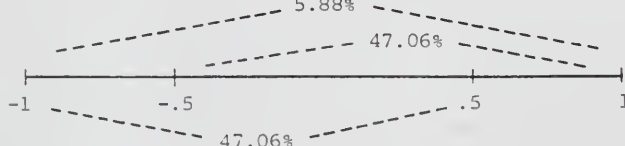
By taking in conjunction the information given by Tables 4.10 and 4.11, the following recommendations are given. If no a priori information about β_1 is available, or if it is believed that $|\beta_1| \leq .6$, then Design 5 is a good design to use. If it is believed that $|\beta_1| > .6$, then Design 6 is the better design.

The relative efficiencies were also calculated for the design $\{\mu_{-1,-.5} = \mu_{-.5,0} = \mu_{0,.5} = \mu_{.5,1} = .25\}$. They were in the vicinity of 25% efficient. Hence this design and Designs 1-3, any of which an experimenter could naively elect to use, are in fact poor designs for fitting a quadratic model.

Table 4.10. Efficiency of designs

Design	Condition	Minimum			Maximum			Median		
		D	V	M	D	V	M	D	V	M
	$|\beta_1| \leq .6$	.954	.934	.829	.968	.964	.897	.963	.940	.871
1	$|\beta_1| > .6$	.918	.914	.770	.965	.970	.880	.942	.933	.809
	$|\beta_1| \leq 1$	.918	.914	.770	.968	.970	.897	.962	.940	.860
	$|\beta_1| \leq .6$	.736	.723	.648	.763	.765	.722	.744	.733	.683
2	$|\beta_1| > .6$	.727	.724	.616	.773	.778	.689	.743	.738	.651
	$|\beta_1| \leq 1$	.727	.723	.616	.773	.778	.722	.744	.733	.669
	$|\beta_1| \leq .6$	.616	.604	.548	.650	.649	.608	.626	.615	.571
3	$|\beta_1| > .6$	.623	.615	.532	.668	.672	.582	.636	.630	.543
	$|\beta_1| \leq 1$	.616	.604	.532	.668	.672	.608	.629	.620	.564
	$|\beta_1| \leq .6$	.988	.963	.795	1	.989	.953	.998	.980	.899
4	$|\beta_1| > .6$	.935	.921	.694	.992	.966	.889	.967	.942	.763
	$|\beta_1| \leq 1$	.935	.921	.694	1	.989	.953	.996	.975	.865
	$|\beta_1| \leq .6$	.970	.986	.796	1	1	.993	.994	.998	.936
5	$|\beta_1| > .6$	.966	.947	.692	.998	.981	.943	.988	.967	.790
	$|\beta_1| \leq 1$	.966	.947	.692	1	1	.993	.991	.993	.910
	$|\beta_1| \leq .6$	.905	.945	.772	.993	.993	1	.958	.972	.947
6	$|\beta_1| > .6$	.974	.936	.683	.999	.999	.995	.997	.984	.813
	$|\beta_1| \leq 1$	.905	.936	.683	.999	.999	1	.972	.974	.917
	$|\beta_1| \leq .6$	.891	.798	.645	.961	.891	.828	.931	.840	.746
7	$|\beta_1| > .6$	.798	.717	.530	.907	.797	.677	.853	.754	.591
	$|\beta_1| \leq 1$	.798	.717	.530	.961	.891	.828	.920	.825	.702

Table 4.11. Bias of designs

Design	Mean ranking Bias	J

1. 7.00 6.00

2. -1   -1/3   1/3   1   5.94 5.59
Each of 6 pairs compared an equal number of times

3. -1   -.5   0   .5   1   2.41 5.97
Each of 10 pairs compared an equal number of times

4. 40% 40% 20% 5.06 3.43

5. 45% 45% 10% 3.85 2.18

6. 50% 50% 2.74 1.65

7. 5.88% 47.06% 47.06% 1.00 3.19

CHAPTER 5

DESIGNS THAT PROTECT AGAINST BIAS

5.1 Introduction

This chapter is concerned with finding designs which protect against the bias present when the true model is cubic, and a quadratic model is fit. Section 5.2 contains a discussion of designs which minimize the bias. Section 5.3 presents designs which minimize J, the integrated mean square error, where

$$J = \int_{-1}^{1} E(\ln \hat{\pi}_x - g(x))^2 \, dx$$

$$= \int_{-1}^{1} E(\ln \hat{\pi}_x - E(\ln \hat{\pi}_x))^2 \, dx + \int_{-1}^{1} (E(\ln \hat{\pi}_x) - g(x))^2 \, dx$$

$$= V + B, \tag{5.1.1}$$

and where

$$g(x) = \beta_1^* x + \beta_2^* x^2 + \beta_3^* x^3, \tag{5.1.2}$$

$$\ln \hat{\pi}_x = \hat{\beta}_1 x + \hat{\beta}_2 x^2, \tag{5.1.3}$$

$$\ln \pi_x = \beta_1 x + \beta_2 x^2. \tag{5.1.4}$$

In the present chapter, β_1^*, β_2^*, and β_3^* represent the true parameter values. Notice that designs which minimize V in (5.1.1) when $\beta_3^*=0$ were found in Section 4.3. Since $\beta_3^*=0$ implies B=0, these designs also minimize J. An expression for V is given by (4.3.3). However the λ_{ij} in the expression are now a function of β_1^*, β_2^*, and β_3^* rather than β_1 and β_2. A general expression for λ_{ij} is found in (2.3.10).

5.2 All Bias Designs

From (5.1.2) and (5.1.3),

$$(E(\ln \hat{\pi}_x) - g(x))^2 = (\,(\beta_1 - \beta_1^*)x + (\beta_2 - \beta_2^*)x^2 - \beta_3^* x^3\,)^2$$

$$= \begin{bmatrix} (\beta_1 - \beta_1^*)^2 x^2 + 2(\beta_1 - \beta_1^*)(\beta_2 - \beta_2^*)x^3 \\[2mm] + (\,(\beta_2 - \beta_2^*)^2 - 2\beta_3^*(\beta_1 - \beta_1^*)\,)x^4 \\[2mm] - 2\beta_3^*(\beta_2 - \beta_2^*)x^5 + \beta_3^{*2}x^6 \end{bmatrix} \qquad (5.2.1)$$

Then from (5.1.1) and (5.2.1), the bias is

$$B = \int_{-1}^{1} (E(\ln \hat{\pi}_x) - g(x))^2\, dx$$

$$= \frac{2}{3}(\beta_1 - \beta_1^*)^2 + \frac{2}{5}(\beta_2 - \beta_2^*) + \frac{2}{7}\beta_3^{*2} - \frac{4}{5}\beta_3^*(\beta_1 - \beta_1^*) . \qquad (5.2.2)$$

Notice that any design which results with the values of β_1 and β_2 that minimizes (5.2.2) is a minimum bias design. The optimal values of β_1 and β_2 as functions of 3_1^*, β_2^*, and β_3^* are now found. Taking derivatives in (5.2.2),

$$\frac{\partial B}{\partial \beta_1} = \frac{4}{3}(\beta_1 - \beta_1^*) - \frac{4}{5}\beta_3^* ,$$

(5.2.3)

$$\frac{\partial B}{\partial \beta_2} = \frac{2}{5}(\beta_2 - \beta_2^*) .$$

Setting the expressions in (5.2.3) equal to zero results in

$$\beta_1 = \frac{3}{5}\beta_3^* + \beta_1^* ,$$

(5.2.4)

$$\beta_2 = \beta_2^* .$$

The three roots to the equation formed by equating (5.1.2) and (5.1.4), after substituting (5.2.4) into (5.1.4), are x=-.7746,0,.7746. These three roots are the points at which the quadratic and cubic curves intersect.

The expected number of times that x_i is preferred over x_j is defined as $n_{ij}P(x_i \rightarrow x_j)$, where the probability that x_i is preferred over x_j is calculated using the true cubic model. This is referred to as "expected" data. If the only levels compared are -.7746, 0, and .7746, then the "expected" data under the true cubic model is the same as the "expected" data under the quadratic model given by (5.2.4). Hence the maximum likelihood procedure using "expected" data under the true model results in β_1 and β_2 when only the three levels -.7746, 0, and .7746 are pairwise compared. This procedure is equivalent to first finding maximum likelihood estimates, $\hat{\beta}_1$ and $\hat{\beta}_2$, from the data, and then taking expectations. This was verified for the present situation.

Therefore, any paired comparisons among the levels x=-.7746,0,.7746 results in $E\hat{\beta}_1=\beta_1$ and $E\hat{\beta}_2=\beta_2$, where β_1 and β_2 are given in (5.2.4), and

hence any such design minimizes the bias. Notice that the optimal design does not depend on the parameters.

This choice of equally optimal designs does not appear to be reasonable, and so the designs are not recommended to be used in practice. Minimizing the bias could result in a large average variance, and hence a large integrated mean square error, J. The next section is a consideration of designs which minimize J.

5.3 Integrated Mean Square Error Designs

Define the function f as

$$f(\beta_1,\beta_2,\beta_3,x) \;=\; \ln \pi_x \;=\; \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \;. \tag{5.3.1}$$

Then it is clear that

$$
\begin{aligned}
f(\beta_1,\beta_2,\beta_3,x) \;&=\; -f(-\beta_1,-\beta_2,-\beta_3,x)\\[2mm]
&=\; -f(\beta_1,-\beta_2,\beta_3,-x)\\[2mm]
&=\; f(-\beta_1,\beta_2,-\beta_3,-x) \;,
\end{aligned}
\tag{5.3.2}
$$

and

$$
\begin{aligned}
f(\beta_1,\beta_2,-\beta_3,x) \;&=\; -f(-\beta_1,-\beta_2,\beta_3,x)\\[2mm]
&=\; -f(\beta_1,-\beta_2,-\beta_3,-x)\\[2mm]
&=\; f(-\beta_1,\beta_2,\beta_3,-x) \;.
\end{aligned}
\tag{5.3.3}
$$

Intuitively, this would imply that optimal designs only need to be found for $\beta_1 \geq 0$ and $\beta_2 \geq 0$ since f is just transposed about the x-axis or y-axis, or both, for the case of negative values of β_1 or β_2. For

example, by (5.3.2) the function f for $(\beta_1, -\beta_2, \beta_3)$ if $f(\beta_1, \beta_2, \beta_3, x)$

transposed about both the x-axis and y-axis. From (3.2.5), ϕ_{ij} can be

written

$$\phi_{ij} \; = \; \frac{\exp(f(\beta_1, \beta_2, \beta_3, x_i) \; + \; f(\beta_1, \beta_2, \beta_3, x_j))}{(\; \exp(f(\beta_1, \beta_2, \beta_3, x_i)) \; + \; \exp(f(\beta_1, \beta_2, \beta_3, x_j)) \;)^2} \; . \; (5.3.4)$$

The fact that only $\beta_1 \geq 0$ and $\beta_2 \geq 0$ need to be considered can be veri-

fied by using (5.3.4) in a manner similar to the proof of Theorem 4.1.

A Fortran program performed a grid search for the designs which

minimize J, where J is given by (5.1.1). The total number of compari-

sons was taken to be N=100, and so V, as given by (4.3.3), was divided

by 100 before substituting into (5.1.1). The grid search was performed

for combinations of $\beta_1, \beta_2 = 0, .5, 1$, and $\beta_3 = -1(.5)1$, plus a few additional

parameter combinations. The levels were fixed at x=-1,-.5,0,.5,1. The

grid allowed for all 10 possible paired comparisons. The search ended

when all 10 comparison proportions were accurate to at least ±.01.

These optimal designs are presented in Table 5.1. The value of J

is also given. As was previously discussed, the designs for negative

β_1 or β_2 are found as follows. First of all, the design for $|\beta_1|$,

$|\beta_2|$, and β_3 is located in the table. If β_1 and β_2 are both positive

or both negative, then the design is found directly from Table 5.1. If

this is not the case, then n_{x_i, x_j} becomes $n_{-x_i, -x_j}$ for all pairs in the

design.

Notice that the pairs (-1,-.5), (-.5,0), and (0,.5) are never com-

pared, and (.5,1) is only compared when $\beta_1 = \beta_2 = \beta_3 = 1$. The designs pre-

sented are not intended to be precise, but rather they give a rough

description of designs which protect against bias. For example, the designs should be symmetric when $\beta_2=0$, but Table 5.1 shows them to be close to but not exactly symmetric.

Because of the restriction of the five fixed levels, the minimum value of J found in the present section is not necessarily the minimum for the class of all possible designs. In fact, when $\beta_3=0$, Designs 4-6 in Section 4.5 have slightly smaller values of J than those given in Table 5.1. This can be seen in Table 5.2 which compares the values of J for the optimal designs presently being discussed with Designs 4-6. This table shows that if the cubic parameter is relatively large, then the present designs are significantly better for protection against bias. Otherwise Designs 4-6 adequately protect against bias, and they additionally are highly efficient for the optimal properties discussed in Chapter 4.

Table 5.1. J-criterion designs

β_1	β_2	β_3	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,.5}$	$\mu_{-.5,1}$	$\mu_{-.5,.5}$	$\mu_{-1,1}$	$\mu_{.5,1}$	J^*
0	0	-1	.13	.12	.26	.26	.24	0	0	.1128
0	0	-.5	.11	.11	.39	.39	0	0	0	.0574
0	0	0	.39	.37	0	.01	0	.23	0	.0366
0	0	.5	.11	.11	.39	.39	0	0	0	.0574
0	0	1	.13	.12	.26	.26	.24	0	0	.1128
0	.5	-1	.19	.15	.21	.17	.29	0	0	.1146
0	.5	-.5	.07	.14	.44	.35	0	0	0	.0590
0	.5	0	.39	.36	.01	.03	0	.22	0	.0383
0	.5	.5	.14	.07	.35	.44	0	0	0	.0590
0	.5	1	.15	.19	.17	.21	.29	0	0	.1146
0	1	-1	.09	.09	.41	.25	.16	0	0	.1212
0	1	-.5	.03	.14	.49	.33	0	0	0	.0641
0	1	0	.38	.36	.01	.02	0	.23	0	.0439
0	1	.5	.14	.03	.33	.49	0	0	0	.0641
0	1	1	.09	.09	.25	.41	.16	0	0	.1212
.5	0	-1	.05	.05	.37	.36	.17	0	0	.1020
.5	0	-.5	.10	.11	.40	.39	0	0	0	.0547
.5	0	0	.39	.38	.01	.01	0	.22	0	.0405
.5	0	.5	.17	.17	.33	.33	0	0	0	.0706
.5	0	1	.30	.29	.01	.01	.39	0	0	.1310
.5	.5	-1	0	.02	.46	.42	.10	0	0	.1034
.5	.5	-.5	.08	.12	.41	.39	0	0	0	.0563
.5	.5	0	.38	.39	0	.01	0	.23	0	.0424
.5	.5	.5	.18	.14	.28	.41	0	0	0	.0728
.5	.5	1	.24	.36	.01	.01	.39	0	0	.1342

* N=100, true model cubic

Table 5.1 - continued

β_1	β_2	β_3	$\mu_{-1,0}$	$\mu_{0,1}$	$\mu_{-1,.5}$	$\mu_{-.5,1}$	$\mu_{-.5,.5}$	$\mu_{-1,1}$	$\mu_{.5,1}$	J^*
.5	1	-1	0	.03	.47	.39	.11	0	0	.1079
.5	1	-.5	.04	.13	.44	.40	0	0	0	.0611
.5	1	0	.39	.29	0	.11	0	.21	0	.0484
.5	1	.5	.20	.06	.23	.52	0	0	0	.0795
.5	1	1	.18	.43	.06	0	.34	0	0	.1439
1	0	-1	.08	.08	.32	.32	.20	0	0	.1021
1	0	-.5	.15	.16	.35	.34	0	0	0	.0615
1	0	0	.36	.43	.09	.02	0	.11	0	.0528
1	0	.5	.33	.34	.06	.03	.24	0	0	.0955
1	0	1	.31	.31	0	0	.38	0	0	.1613
1	.5	-1	.06	.08	.34	.34	.18	0	0	.1036
1	.5	-.5	.13	.18	.34	.35	0	0	0	.0634
1	.5	0	.18	.54	.28	0	0	0	0	.0548
1	.5	.5	.22	.43	.16	.01	.19	0	0	.0986
1	.5	1	.26	.38	0	0	.37	0	0	.1657
1	1	-1	0	.03	.42	.46	.09	0	0	.1082
1	1	-.5	.11	.20	.33	.37	0	0	0	.0694
1	1	0	.13	.59	.28	0	0	0	0	.0631
1	1	.5	.20	.48	.13	0	.19	0	0	.1085
1	1	1	.22	.27	0	0	.38	0	.13	.1786
.4	.4	-1	.09	.09	.32	.29	.22	0	0	.1043
.4	.6	.2	.29	.13	.19	.40	0	0	0	.0480
.8	.6	-1	.04	.05	.38	.38	.16	0	0	.1028

* N=100, true model cubic

Table 5.2. Values of J

β_1	β_2	β_3	Design 4	Design 5	Design 6	"Minimum" J
0	0	-1	.1433	.1327	.1299	.1128
0	0	-.5	.0638	.0601	.0592	.0574
0	0	0	.0357	.0350	.0359	.0366
0	0	.5	.0638	.0601	.0592	.0574
0	0	1	.1433	.1327	.1299	.1128
0	.5	-1	.1470	.1361	.1332	.1146
0	.5	-.5	.0659	.0622	.0616	.0590
0	.5	0	.0374	.0367	.0380	.0383
0	.5	.5	.0659	.0622	.0616	.0590
0	.5	1	.1470	.1361	.1332	.1146
0	1	-1	.1582	.1467	.1438	.1212
0	1	-.5	.0725	.0690	.0693	.0641
0	1	0	.0428	.0424	.0448	.0439
0	1	.5	.0725	.0690	.0693	.0641
0	1	1	.1582	.1467	.1438	.1212
.5	0	-1	.1374	.1259	.1215	.1020
.5	0	-.5	.0612	.0576	.0570	.0547
.5	0	0	.0400	.0390	.0392	.0405
.5	0	.5	.0754	.0709	.0690	.0706
.5	0	1	.1626	.1516	.1486	.1310
.5	.5	-1	.1402	.1284	.1238	.1034
.5	.5	-.5	.0631	.0596	.0591	.0563
.5	.5	0	.0419	.0410	.0417	.0424
.5	.5	.5	.0784	.0741	.0725	.0728
.5	.5	1	.1683	.1572	.1541	.1342

Note: Designs 4-6 defined in Section 4.5, N=100, true model cubic

Table 5.2 - continued

β_1	β_2	β_3	Design 4	Design 5	Design 6	"Minimum" J
.5	1	-1	.1489	.1364	.1312	.1079
.5	1	-.5	.0692	.0657	.0660	.0611
.5	1	0	.0479	.0476	.0498	.0484
.5	1	.5	.0876	.0840	.0839	.0795
.5	1	1	.1858	.1745	.1718	.1439
1	0	-1	.1377	.1257	.1203	.1021
1	0	-.5	.0661	.0625	.0614	.0615
1	0	0	.0547	.0523	.0504	.0528
1	0	.5	.1015	.0948	.0899	.0955
1	0	1	.2060	.1919	.1846	.1613
1	.5	-1	.1404	.1281	.1225	.1036
1	.5	-.5	.0683	.0648	.0641	.0634
1	.5	0	.0576	.0556	.0542	.0548
1	.5	.5	.1069	.1005	.0959	.0986
1	.5	1	.2162	.2019	.1943	.1657
1	1	-1	.1486	.1357	.1294	.1082
1	1	-.5	.0752	.0721	.0726	.0694
1	1	0	.0664	.0658	.0666	.0631
1	1	.5	.1232	.1182	.1152	.1085
1	1	1	.2477	.2333	.2257	.1786

Note: Designs 4-6 defined in Section 4.5, N=100, true model cubic

CHAPTER 6

DESIGNS FOR PRELIMINARY TEST ESTIMATORS

6.1   Introduction

The motivation for this chapter is a paper by Sen (1979) on pre-
liminary test maximum likelihood estimators (PTMLE). The procedure for
the PTMLE is to collect the entire data and then make a test of hypoth-
esis. If the null hypothesis is rejected, then the maximum likelihood
estimates under an unrestricted parameter space are found. If the null
hypothesis is accepted, the estimates are found under a restricted pa-
rameter space.

Presently, the unrestricted parameter space is the two-dimensional
(β_1, β_2), the restricted parameter space is $(\beta_1, 0)$, and hence the null
hypothesis is H_o: $\beta_2 = 0$. The objective of the present chapter is to
find designs which minimize the integrated mean square error using the
preliminary test estimator. This is similar to Section 4.3, except
that in the present situation a linear model may be fit to the data
instead of simply fitting a quadratic model regardless of the data.

Define $(\hat{\beta}_1, 0)$ as the maximum likelihood estimator under the re-
stricted parameter space, $(\tilde{\beta}_1, \tilde{\beta}_2)$ as the estimator under the unre-
stricted parameter space, and $(\hat{\beta}_1^*, \hat{\beta}_2^*)$ as the PTMLE. Sen shows that for
any fixed alternative, i.e. for any $\beta_2 \neq 0$, $\sqrt{N}(\tilde{\beta}_1 - \beta_1, \tilde{\beta}_2 - \beta_2)$ and
$\sqrt{N}(\hat{\beta}_1^* - \beta_1, \hat{\beta}_2^* - \beta_2)$ have the same asymptotic multi-normal distribu-
tion, and therefore the optimal designs for the PTMLE are the same as

145

the optimal designs found in Section 4.3. Likewise, the optimal designs for $\beta_2=0$ are identical to the ones presented in Section 3.2.

The remaining situation is when (β_1,β_2) is "close" to the restricted parameter space, i.e. when β_2 is "close" to zero. This is conceived by defining a sequence $\{K_N\}_{N=1}^{\infty}$ of local alternatives to be

$$K_N: \quad \beta_2 = \gamma/\sqrt{N}, \tag{6.1.1}$$

where γ is some scalar constant. Notice that as $N\to\infty$, $K_N \to H_o$. This situation is covered in the next section.

6.2 Average-variance Designs for Preliminary
Test Maximum Likelihood Estimators

For the remainder of the present chapter, the sequence of alternatives given in (6.1.1) is considered. The asymptotic dispersion matrix for the PTMLE in general is given by (5.8) on page 1029 of Sen's paper. After working through Sen's paper to find this matrix for the present situation, the asymptotic dispersion matrix for $\sqrt{N}(\hat{\beta}_1^* - \beta_1, \hat{\beta}_2^* - \beta_2)$ is

$$\Sigma = \left[\begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{pmatrix}^{-1} + \left[\begin{pmatrix} 1/\lambda_{11} & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{pmatrix}^{-1} \right] P(\chi^2_{3,\Delta} \leq \chi^2_{1,\alpha}) \right.$$
$$\left. + \underline{\gamma}^* \underline{\gamma}^{*'} (2P(\chi^2_{3,\Delta} \leq \chi^2_{1,\alpha}) - P(\chi^2_{5,\Delta} \leq \chi^2_{1,\alpha})) \right],$$

where λ_{11}, λ_{12}, and λ_{22} are given by (3.3.2)-(3.3.4), $\chi^2_{r,\Delta}$ is a non-central chi-square random variable with noncentrality parameter Δ and r degrees of freedom, $\chi^2_{1,\alpha}$ is such that $P(\chi^2_{1,\Delta=0} > \chi^2_{1,\alpha}) = \alpha$, and

$$\Delta = \gamma^2 \frac{\lambda_{11}\lambda_{22} - \lambda_{12}^2}{\lambda_{11}} , \tag{6.2.2}$$

$$\gamma^* = \gamma \begin{pmatrix} -\lambda_{12}/\lambda_{11} \\ 1 \end{pmatrix} . \tag{6.2.3}$$

Notice that in (6.2.1) noncentral chi-square probabilities need to be found. A number of convenient approximations exist. One such approximation is to use a central chi-square probability to approximate a noncentral chi-square probability. This is found in the Handbook of Mathematical Functions, published by the National Bureau of Standards. Letting ν denote the degrees of freedom and Δ denote the noncentrality parameter, $P(\chi^2_{\nu,\Delta} \le c)$ can be approximated by $P(\chi^2_{\nu*,0} \le c/(1+b))$, where $b = \Delta/(\nu + \Delta)$, and $\nu* = (\nu + \Delta)/(1 + b)$.

Haynam, Govindarajulu, and Leone (1962) present a computational formula for noncentral chi-square probabilities. The formula is

$$P(\chi^2_{\nu,\lambda} > y) = \sum_{j=0}^{\infty} \frac{e^{-\lambda/2}(\lambda/2)^j}{j!} \left[\frac{1}{\Gamma((\nu/2)+j)} \int_y^{\infty} (x/2)^{(\nu/2)+j-1} e^{-x/2} \, d(x/2) \right]$$

$$= \sum_{j=0}^{\infty} \frac{e^{-\lambda/2}(\lambda/2)^j}{j!} P(\chi^2_{\frac{\nu}{2}+j} > y) . \tag{6.2.4}$$

The value of $j!$ becomes large very rapidly, and hence the summation in (6.2.4) also converges rapidly.

These two methods were compared to observe how close the approximation is to the actual probability, since the computing costs would be reduced if the approximation was adequate. This comparison was made

for combinations of $\nu=3,5$, $\Delta=.05(.05)2$, and $\alpha=.05(.05).30$. These values were chosen because they are the type of probabilities that need to be calculated in (6.2.1). The summation in (6.2.4) was terminated when the jth term was less than 10^{-6}. This turned out to be between 5 and 11 terms in all cases considered.

For small α, Δ, and $\nu=3$, the approximation formula slightly over-estimates the probability. However, for $\nu=3$, $\alpha \geq .20$, and large values of Δ, the approximation underestimates the probability by as much as .01. For this reason, it was decided to use the computational formula (6.2.3) in the discussion which follows.

Once again, a Fortran program was written to conduct a grid search for optimal designs. This was done for $N=100$ total comparisons and all combinations of $\beta_1, \beta_2 = 0(.2)1$. The entire procedure was repeated for error rates $\alpha=.01,.05,.10,.15,.20$. The optimal design minimizes

$$V = \int_{-1}^{1} (\, x \quad x^2\,) \, \Sigma \begin{pmatrix} x \\ x^2 \end{pmatrix} \, dx \, , \tag{6.2.5}$$

where Σ is given in (6.2.1). By (6.1.1), $\gamma=\sqrt{N}\beta_2$ was substituted for γ in (6.2.5). The procedure used for each value of α is identical to the one described in the last subsection of Section 4.2, except that the values of x_1, x_2, and the comparison proportions were found accurate to only $\pm.001$.

The result of the grid search was that for $\beta_2 \leq .2$, the level $\alpha=.01$ gave the smallest value for V. This implies that it may be best to take $\alpha=0$. This further implies that a linear model should be fit regardless of the data. For $\beta_2 \geq .4$, it appeared that it is best to

use $\alpha=1$, implying a quadratic model should be fit regardless of the data, and hence the optimal designs are found in Section 4.3.

This is a very peculiar result. It seems to be implying that there is some value of β_2, say β_{20}, in the interval $(.2,.4)$ such that a linear model should be fit for any $|\beta_2|$ less than β_{20}, and a quadratic model should be fit for any $|\beta_2|$ greater than β_{20}. No definite explanation for this phenomenon is presently known.

CHAPTER 7

TWO-STAGE SAMPLING FOR MODEL FITTING

7.1 Introduction

Suppose the entire experiment is conducted in two stages. The total number of paired comparisons, N, is then split into two parts. The first sampling stage has N_1 total comparisons, and the second has N_2 total comparisons, such that $N_1+N_2=N$. After the first stage, the hypothesis H_o: $\beta_2=0$ is tested. If the hypothesis is rejected, then the design of the second stage would be one which is optimal for fitting a quadratic model. The quadratic model is then fit using only the data from the second stage. If the hypothesis is accepted, then the second stage would be a design optimal for fitting a linear model, again using only the second stage data for the estimation of the linear model.

The first stage design, referred to as Design 1, is $\{n_{-1,0} = n_{0,1} = .5N_1\}$. This design was shown in Section 3.3 to minimize $\text{var}(\hat{\beta}_2)$ for $\beta_1=\beta_2=0$. The design was also shown to be highly efficient for $|\beta_1| \le 1$. This is a good choice for the first stage design because it maximizes the power of the test of the null hypothesis for a given value of N_1.

The second stage design is one of two choices depending on whether or not it is decided to fit a quadratic model. If the null hypothesis is rejected, the design for the second stage, referred to as Design 2Q, is $\{n_{-.24,1} = n_{-1,.24} = .4N_2,\ n_{-1,1} = .2N_2\}$. This design was shown in

150

Section 4.5 to be one of the overall better designs for fitting a quad-

ratic model. Finally, if the null hypothesis is accepted, the design

$\{n_{-1,1} = N_2\}$ is to be run. This design was shown in Section 3.2 to be

optimal for minimizing $\text{var}(\hat{\beta}_1)$, and hence optimal for fitting a linear

model. This design is referred to as Design 2L. The following section

discusses optimal choices of the error rate for the test of hypothesis

and the optimal values of N_1 and N_2 using a Bayesian approach.

7.2 Optimal Error Rate

Two prior bivariate distributions for (β_1, β_2) and two optimality

criteria are considered using a Bayesian approach. The prior distribu-

tions are the uniform and normal distributions, both centered about the

point $(0,0)$. The uniform prior is appropriate if an experimenter could

state that β_1 and β_2 are in some intervals centered about zero, but

with no reason to believe that some subset of values for β_1 and β_2 is

more likely than any other. On the other hand, if the person could

state that β_1 and β_2 are in some intervals centered about zero, and

furthermore that it is more likely that they are near zero than it is

that they are far away from zero, then the normal prior is appropriate.

The optimality criteria considered are the minimization of the

integrated mean square error and D-optimality. These two criteria have

been previously discussed in Chapter 4 for the case of no bias, and are

also applicable for the present situation. The difference between the

situation in Chapter 4 and the situation in the present chapter is that

bias can appear in the present situation. That is, there is bias pre-

sent if a linear model is fit when the true model is quadratic. For

this reason, the mean square error of the maximum likelihood estimators

is used in the present chapter instead of the variance-covariance
matrix.

The decision rule, d, is a choice of either the linear design (1)
or the quadratic design (q) for the second stage of the experiment.
Notice that the decision rule is equivalent to choosing the error rate
for the test of H_o: $\beta_2 = 0$. For the present section, the Bayes rule is
found for fixed N_1 and N_2, such that $N_1 + N_2 = 100$.

Define p to be the probability of rejecting the null hypothesis.
The value of p depends on β_2, α, and $\text{var}(\hat{\beta}_2)$ from the first stage. In
fact, it can be shown that

$$p \quad = \quad 1 - \Phi(z_2) + \Phi(z_1) \; , \qquad (7.2.1)$$

where

$$z_1 \quad = \quad -z_{\alpha/2} - \beta_2/(\text{var}(\hat{\beta}_2))^{.5} \; , \qquad (7.2.2)$$

$$z_2 \quad = \quad z_{\alpha/2} - \beta_2/(\text{var}(\hat{\beta}_2))^{.5} \; , \qquad (7.2.3)$$

and where Φ is the standard normal cumulative distribution function,
and $z_{\alpha/2}$ is such that $(1 - \Phi(z_\alpha)) = \alpha$. Because of the symmetry of the
first stage design, by (3.3.9) the standard error of $\hat{\beta}_2$ is

$$(\text{var}(\hat{\beta}_2))^{.5} \quad = \quad 1/\sqrt{\lambda_{22}} \quad = \quad ((N_1 \phi_{-1,0}/2) + (N_1 \phi_{0,1}/2))^{-.5} \; , \quad (7.2.4)$$

where $\phi_{-1,0}$ and $\phi_{0,1}$ are given in (4.2.9).

Consider first the integrated mean square error criterion. The
loss function for this criterion is

$$L(\underline{\beta},1) \quad = \quad \int_{-1}^{1} \text{MSE}(\ln \hat{\pi}_x) \; dx \; , \qquad (7.2.5)$$

$$L(\underline{\beta}, q) \quad = \quad \int_{-1}^{1} \text{MSE}(\ln \tilde{\pi}_x) \, dx \ , \qquad (7.2.6)$$

where

$$\underline{\beta} \quad = \quad (\beta_1, \beta_2) \ ,$$

$$\ln \hat{\pi}_x \quad = \quad \hat{\beta}_1 x \ , \qquad (7.2.7)$$

$$\ln \tilde{\pi}_x \quad = \quad \tilde{\beta}_1 x + \tilde{\beta}_2 x^2 \ , \qquad (7.2.8)$$

and where $\hat{\beta}_1$ is the maximum likelihood estimator from Design 2L, and $(\tilde{\beta}_1, \tilde{\beta}_2)$ is the maximum likelihood estimator from Design 2Q. The risk function is then

$$R(\underline{\beta}, d) \quad = \quad (1 - p) L(\underline{\beta}, 1) + p L(\underline{\beta}, q) \ . \qquad (7.2.9)$$

The mean square error of $(\hat{\beta}_1, 0)$ is

$$\text{MSE}(\hat{\beta}_1, 0) \quad = \quad \begin{pmatrix} 1/\lambda_{11}{}^* & 0 \\ & \\ 0 & \beta_2{}^2 \end{pmatrix} \ , \qquad (7.2.10)$$

where $\phi_{-1,1}$ is given in (4.2.9), and

$$\lambda_{11}{}^* \quad = \quad 4N_2 \phi_{-1,1} \ , \qquad (7.2.11)$$

From (7.2.5) and (7.2.10), the loss when Design 2L is chosen is

$$L(\underline{\beta}, 1) \quad = \quad \int_{-1}^{1} \begin{pmatrix} x & x^2 \end{pmatrix} \begin{pmatrix} 1/\lambda_{11}{}^* & 0 \\ & \\ 0 & \beta_2{}^2 \end{pmatrix} \begin{pmatrix} x \\ \\ x^2 \end{pmatrix} dx$$

$$= \quad \frac{2/3}{\lambda_{11}{}^*} + \frac{2}{5} \beta_2{}^2 \ . \qquad (7.2.12)$$

The mean square error of $(\tilde{\beta}_1, \tilde{\beta}_2)$ is

$$
\text{MSE}(\tilde{\beta}_1, \tilde{\beta}_2) \;=\; \begin{pmatrix} 1/\lambda_{11} & 0 \\[2mm] 0 & 1/\lambda_{22} \end{pmatrix} \;, \tag{7.2.13}
$$

where from (3.3.2)-(3.3.5),

$$
\lambda_{11} \;=\; N_2(.615(\phi_{-1,.24} + \phi_{-.24,1}) + .8\phi_{-1,1}) \;, \tag{7.2.14}
$$

$$
\lambda_{22} \;=\; N_2(.377(\phi_{-1,.24} + \phi_{-.24,1})) \;, \tag{7.2.15}
$$

$$
\phi_{-1,.24} \;=\; \frac{e^{-.76\beta_1 + 1.0576\beta_2}}{(e^{\beta_2 - \beta_1} + e^{.24\beta_1 + .0576\beta_2})^2} \;, \tag{7.2.16}
$$

$$
\phi_{-.24,1} \;=\; \frac{e^{.76\beta_1 + 1.0576\beta_2}}{(e^{-.24\beta_1 + .0576\beta_2} + e^{\beta_1 + \beta_2})^2} \;, \tag{7.2.17}
$$

and $\phi_{-1,1}$ is given in (4.3.9). The off-diagonal elements of (7.2.10) and (7.2.13) are zero because both second stage designs are symmetric. From (7.2.6) and (7.2.13), the loss when Design 2Q is chosen is

$$
L(\underline{\beta}, q) \;=\; \int_{-1}^{1} \begin{pmatrix} x & x^2 \end{pmatrix} \begin{pmatrix} 1/\lambda_{11} & 0 \\[2mm] 0 & 1/\lambda_{22} \end{pmatrix} \begin{pmatrix} x \\[2mm] x^2 \end{pmatrix} dx
$$

$$
\;=\; \frac{2/3}{\lambda_{11}} + \frac{2/5}{\lambda_{22}} \;. \tag{7.2.18}
$$

Then for the integrated mean square error criterion, by (7.2.9), (7.2.12), and (7.2.18) the risk function becomes

$$R(\underline{\beta},d) \quad = \quad (1 - p)\left[\frac{2/3}{\lambda_{11}{}^{\star}} + \frac{2}{5}\beta_2{}^2\right] \quad + \quad p\left[\frac{2/3}{\lambda_{11}} + \frac{2/5}{\lambda_{22}}\right] \quad . \tag{7.2.19}$$

The Bayes risk is defined to be

$$r(\tau,d) \quad = \quad \int R(\underline{\beta},d) \ d\tau(\underline{\beta}) \quad . \tag{7.2.20}$$

For the D-optimal criterion, the loss function is

$$L(\underline{\beta},1) \quad = \quad \begin{vmatrix} 1/\lambda_{11}{}^{\star} & 0 \\ \\ 0 & \beta_2{}^2 \end{vmatrix} \quad = \quad \beta_2{}^2/\lambda_{11}{}^{\star} \quad , \tag{7.2.21}$$

$$L(\underline{\beta},q) \quad = \quad \begin{vmatrix} 1/\lambda_{11} & 0 \\ \\ 0 & 1/\lambda_{22} \end{vmatrix} \quad = \quad 1/\lambda_{11}\lambda_{22} \quad . \tag{7.2.22}$$

Hence the risk function in this case is

$$R(\underline{\beta},d) \quad = \quad (1 - p)\beta_2{}^2/\lambda_{11}{}^{\star} \quad + \quad p/\lambda_{11}\lambda_{22} \quad . \tag{7.2.23}$$

The two prior distributions are now defined. The first prior is the bivariate uniform distribution, whose density function is

$$d\tau(\underline{\beta}) \quad = \quad \begin{cases} 1/4U & \text{if } \beta_1 \epsilon[-1,1], \ \beta_2 \epsilon[-U,U] \ , \\ \\ 0 & \text{otherwise,} \end{cases} \tag{7.2.24}$$

where U is the upper limit to the prior distribution of β_2.

The other prior considered is the bivariate normal distribution. Since the support of the normal distribution is infinite, the density was truncated at three standard deviations about the mean in order to perform the numerical integration. The marginal densities are then

$$
d\tau_{\beta_1}(\beta_1) \;=\;
\begin{cases}
e^{-4.5\beta_1^{\,2}}/.8334 & \text{if } \beta_1 \epsilon [-1,1], \\[2em]
0 & \text{otherwise,}
\end{cases}
$$

$$(7.2.25)$$

$$
d\tau_{\beta_2}(\beta_2) \;=\;
\begin{cases}
e^{-\beta_2^{\,2}/2\sigma^2}/2.5\sigma & \text{if } \beta_2 \epsilon [-U,U], \\[2em]
0 & \text{otherwise,}
\end{cases}
$$

where $\sigma = U/3$. These marginals are the usual normal density divided by .9974 so that the density integrates to 1. The joint prior distribution of (β_1,β_2) is the product of the marginals.

The integral in (7.2.20) can not be analytically found for any criterion and prior since p is a function of the standard normal cumulative distribution, which in turn is a function of $\underline{\beta}$. So a SAS computer program was written to numerically integrate (7.2.20) for the two prior distributions described in the two preceding paragraphs and the two criteria discussed in the present section.

The numerical integration is performed as follows. The rectangle formed by the area of positive density is divided into a 10^2 grid. This results with 100 rectangles of equal size. The expression $R(\underline{\beta},d)d\tau(\underline{\beta})$ is then evaluated at each of the centers of the 100 rectangles. This gives the height of the function $R(\underline{\beta},d)d\tau(\underline{\beta})$ at the center of each rectangle. The lengths of the sides of the rectangles are .02 and .02U, and so the approximate volume of each region formed by the function and the sides of each rectangle is found by multiplying the height of the function at the center of the rectangle by .04U. The approximation of the interval in (7.2.20) is then found by summing these 100 volumes.

For each of the four combinations of the two prior distributions and two criteria, the Bayes risk in (7.2.20) was found for $N_1=10(10)90$, $\alpha=0(.05)1$, and various values of U. The value of N_2 is found by $N_2=100-N_1$. For each value of N_1 and U, the best α level was found from the choices $\alpha=0(.05)1$. These optimal α levels are graphed as a function of U, the upper limit to the prior distibution of β_2, for each of $N_1=10(10)90$. Remember that these graphs are not precise since the optimal α level is only found to the nearest multiple of .05, but they do give a basic understanding for the behavior of the α level as a function of the prior distribution of β_2. The graphs for the integrated mean square error criterion with a uniform prior are found in Figure 7.1, and with a normal prior are found in Figure 7.2. The graphs for the D-optimality criterion with a uniform prior are found in Figure 7.3, and with a normal prior are found in Figure 7.4. The next section contains a discussion of these four figures.

7.3   Concluding Remarks

In Figures 7.1-7.4, the axis labelled "U" corresponds to the upper limit of the prior distribution of β_2. The U-axis in the graphs for the normal prior, Figures 7.2 and 7.4, ranges up to 2.08. The normal prior with U=2 corresponds to β_2 having a normal distribution with mean 0, and standard deviation 2/3. The U-axis of the graphs for the uniform prior, Figures 7.1 and 7.3, ranges up to 1.2. It is likewise true that for the uniform distribution, U=1.2 implies that the standard error of β_2 is also approximately 2/3. So for a given amount of variability, the upper limit for the normal prior is approximately 1.73 times the upper limit for the uniform prior.

For example, if an experimenter believes β_2 is somewhere in the interval (-.5,.5), but has no inclination to where in that interval β_2 is, then a uniform prior would be appropriate. Figure 7.1 indicates that for U=.5 and N_1=10, the test should be run at an error rate of approximately α=.20; for N_1=20, α=.08. A normal prior with the same amount of variability or uncertainty would be found by multiplying U=.5 by 1.73. This results in U=.865. Figure 7.2 indicates that for U=.865 and N_1=10, the test should be run at an error rate of approximately α=.35; for N_1=20, α=.20.

So the error rates for the particular value of the standard error of β_2 discussed in the preceding paragraph appear to be slightly larger for the normal prior than for the uniform prior. However, this is not apparent for the entire range of values for U and N_1. In fact, by looking at the maximum value of U in Figures 7.1 and 7.2, α=1 in both cases for N_1=10,20,30, but α is larger for the uniform prior than for the normal prior when N_1=40(10)80. For most values of the standard error of β_2 and N_1, in particular when the standard error is relatively large, α is larger for the uniform prior than for the normal prior.

Notice from all four figures that for a given value of U and N_1, the α level for the normal prior is smaller than for the uniform prior. This is partially explained by the previous paragraph. The normal prior with the same value of U as a uniform prior will have a smaller standard error, and most of the density is near β_2=0. Consequently, there is a higher probability that β_2 is near zero, and it is therefore reasonable that the hypothesis H_o: β_2=0 should be tested at a smaller α level.

Comparing the two criteria for each of the two priors indicates that for given values of U and N_1, the α level is smaller for the D-optimal criterion than it is for the integrated mean square error criterion.

It turns out that in all four situations, for a given value of U, the Bayes risk is smaller for $N_1=10$ than it is for any of $N_1=20(10)90$. This indicates that if an experimenter is interested in running a two-stage experiment, a relatively small number of comparisons should be used in the first stage, saving the majority of the available comparisons for the second stage. This may be explained by the fact that in the present discussion, the first stage data are not used in the estimation phase. Only the second stage data are used for the final estimation. So the result found at the beginning of the present paragraph implies that it is better to use the large majority of the available comparisons for the sole purpose of testing the null hypothesis.

Further research remains to be conducted on combining the first and second stage data in the estimation phase. This would involve deriving $var(\ln \hat{\pi}_x)$, where $\ln \hat{\pi}_x$ is expressed as

$$\ln \hat{\pi}_x = I(\hat{\beta}_1 x) + (1 - I)(\tilde{\beta}_1 x + \tilde{\beta}_2 x^2) ,$$

and where I is the decision rule for the first stage, i.e.

$$I = \begin{cases} 1 & \text{if it is decided to fit a linear model,} \\ 0 & \text{if it is decided to fit a quadratic model.} \end{cases}$$

This is not a trivial problem because the random variables I, $\hat{\beta}_1$, and

$(\tilde{\beta}_1, \tilde{\beta}_2)$ are not independent. Intuitively, if this could be resolved, it appears reasonable that some optimal ratio N_1/N_2 and α level could be found.
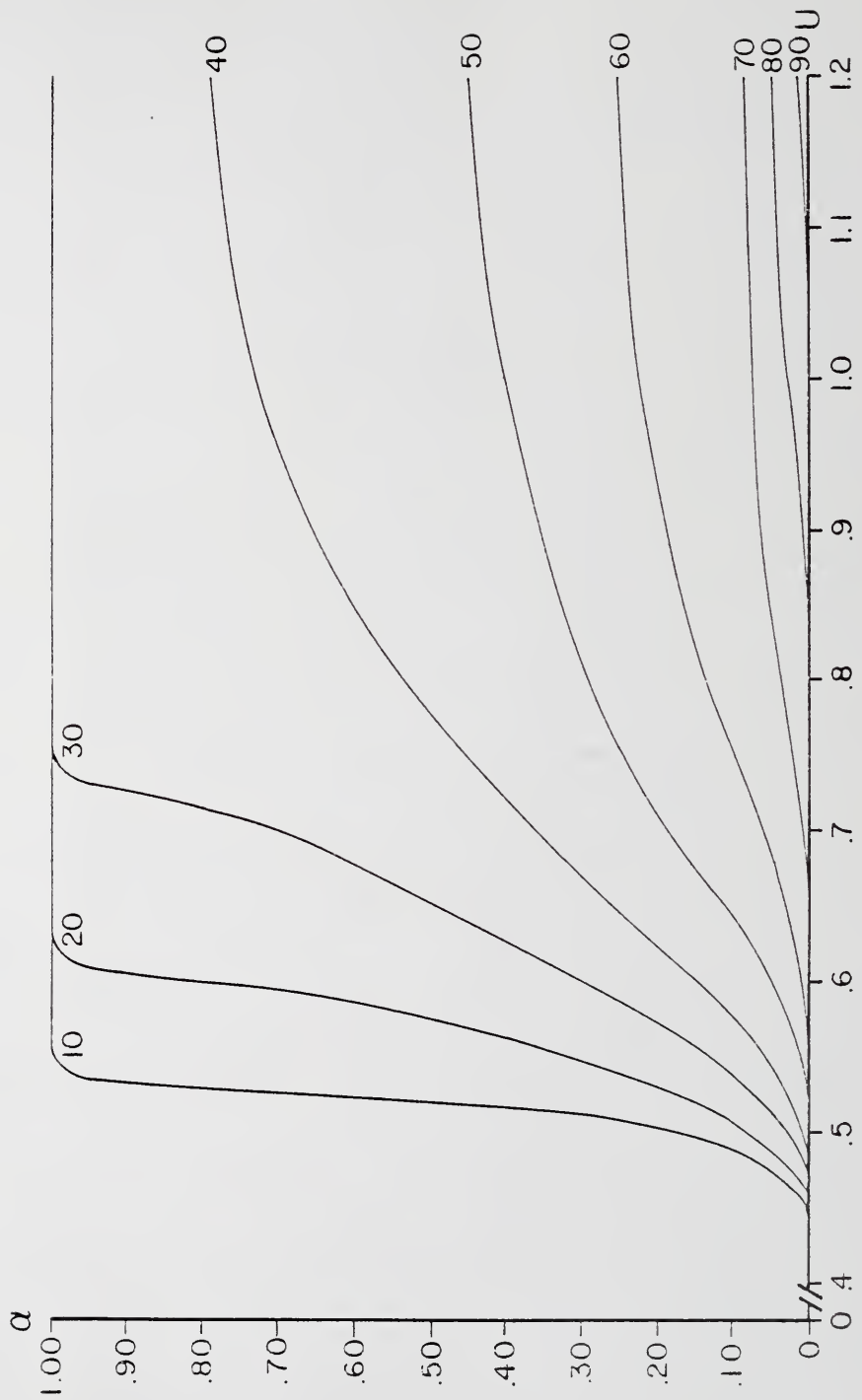
Figure 7.1. Minimization of the average mean square error – Uniform prior

162

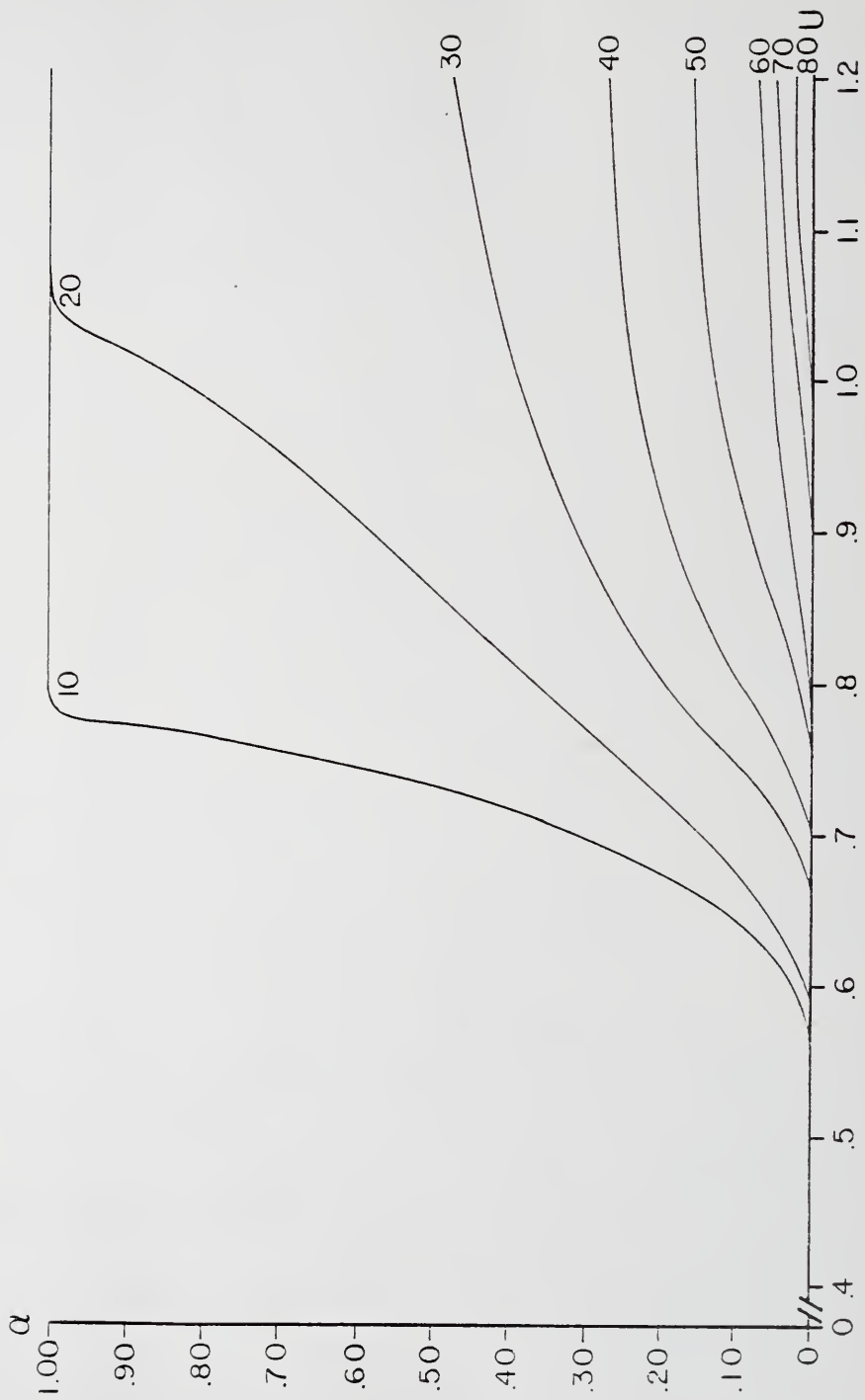Figure 7.2. Minimization of the average mean square error - Normal prior

Figure 7.3. D-optimality - Uniform prior

Figure 7.4. D-optimality - Normal prior

164

APPENDIX A

COMPUTER PROGRAM THAT FINDS
MAXIMUM LIKELIHOOD ESTIMATES OF $\pi_1, \ldots, \pi_t, \theta$

This appendix contains a computer program written in the language APL. There are two matrices which need to be entered onto the computer. First of all, the matrix referred to as "N" is a t×t matrix, with $(i,j)^{th}$ element equal to the number of times treatment T_i is preferred over treatment T_j. The other matrix is also t×t, and is referred to as "TIE". The lower triangular part and the main diagonal of "TIE" are all zeroes. The $(i,j)^{th}$ element, where $i<j$, is the number of ties between T_i and T_j.

The output is the following.

(1) The initial estimates and subsequent iterative estimates until all estimates have sufficiently converged.

(2) The asymptotic variance-covariance matrix of $\sqrt{r}(\hat{\theta} - \theta)$, $\sqrt{r}(\hat{\pi}_1 - \pi_1)$, ... , $\sqrt{r}(\hat{\pi}_t - \pi_t)$, where r is the number of replicates for each pair. (Note that this will not be applicable if some pairs are compared much more often than others.)

(3) The average number of replicates per pair $(\bar{r})$.

(4) The asymptotic variance-covariance matrix of $\hat{\theta}, \hat{\pi}_1$, ... , $\hat{\pi}_t$ under the assumption that each pair of treatments is compared $\bar{r}$ times.

(5) The asymptotic chi-square test of the equality of π_1, ... , π_t. The output includes the chi-square statistic and the degrees of

165

freedom.

If there are no ties, only (1) and (5) above are printed, because in this case the calculation of the element of the inverse to the variance-covariance matrix corresponding to $\hat{\theta}$ results in division by zero.

An iterative procedure is used in the estimation phase of the computer program. The initial estimates of $\theta, \pi_1, \ldots, \pi_t$ are the ones presented by Dykstra (1956). The method used to find the estimates and asymptotic variance-covariance matrix was presented by Rao and Kupper (1967). In their presentation, it was assumed that each pair was compared an equal number of times. Therefore, as pointed out above, the variance-covariance matrices given by the program may not be applicable if the pairs are not compared an equal number of times.

Let $\hat{\theta}^{(k)}, \hat{\pi}_1^{(k)}, \ldots, \hat{\pi}_t^{(k)}$ be the estimates after the kth iteration. The iterative procedure is terminated when

$$\left| \frac{\hat{\theta}^{(k-1)} - \hat{\theta}^{(k)}}{\hat{\theta}^{(k-1)}} \right| \Big/ (1 + t) \; + \; \sum_{i=1}^{t} \left| \frac{\hat{\pi}_i^{(k)} + \hat{\pi}_i^{(k-1)}}{\hat{\pi}_i^{(k-1)}} \right| \; < \; 10^{-4} \; .$$

If so desired, this can be changed simply by altering line [13].

Example. Suppose that

$$N = \begin{bmatrix} 0 & 40 & 20 \\ 0 & 0 & 40 \\ 14 & 0 & 0 \end{bmatrix} \quad , \quad TIE = \begin{bmatrix} 0 & 3 & 5 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \; .$$

Then an execution of the program yields the following:

	θ	π_1	π_2	π_3
Initial estimates	1.0789	.5724	.3236	.1040
Final estimates (19 iterations)	1.1510	.7267	.2091	.0641

Variance-covariance matrix of $\sqrt{r}(\hat{\theta} - \theta), \sqrt{r}(\hat{\pi}_1 - \pi_1), \ldots, \sqrt{r}(\hat{\pi}_t - \pi_t)$:

$$
\begin{bmatrix}
.209696 & .022594 & -.012075 & -.010519 \\
.022594 & .136691 & -.103543 & -.033148 \\
-.012075 & -.103543 & .088575 & .014968 \\
-.010519 & -.033148 & .014968 & .028699
\end{bmatrix}
$$

Average number of replicates: $\overline{r} = 39.5$

Variance-covariance matrix of $\hat{\theta}, \hat{\pi}_1, \ldots, \hat{\pi}_t$ is approximately:

$$
\begin{bmatrix}
.005309 & .000572 & -.000306 & -.000266 \\
.000572 & .003461 & -.002621 & -.000839 \\
-.000306 & -.002621 & .002242 & .000379 \\
-.000266 & -.000839 & .000379 & .000727
\end{bmatrix}
$$

Chi-square test of H_o: $\pi_1 = \cdots = \pi_t$:

$\chi^2 = 57.97$, degrees of freedom = 2.

```
∇ H FAIRCOM TIE;I;BIG;I;SN;LU;ITER;THETA;F;F;FII;IS;GF;CHECK;H;THETA;HF;I;LAMBDA;TEMP;II;
    LAMBDA;SIGMA;F100
[1]   F+BI(+/BR+N+TIE
[2]   LU+(SN+/Q+/U)+(+/Q+/TIE)÷2
[3]   ITER+1+0
[4]   THETA+((2×LU)÷SN)-1
[5]  NEXT2:;
[6]   []+''';[]+'ITER        THETA    TREATMENT PARAMETERS'
[7]   []+'';[]+''
[8]   F0+F÷+/F+(1,T)ρBI÷(LN×(T-1)×2)-BIX((T+F×L1;])-2)
[9]   '13,X2,F10.4,X2,100F7.4' []FMT(ITER,THETA,F0)
[10] TOP2:ITER+ITER+1
[11]  H;THETA+1+(THETA÷THETA+1)×(2×(LU-SN))÷+/R(BR×1÷(Q×BIGF)+THETA×BIGF+(T,T)ρF×II)+,×QF×II
[12]  F+(1,F×II;])ρBI×I+(+/BR÷(Q×BIGF)+THETA×BIGF)+(+/(THETA×Q×BB)÷BIGF+THETA×Q×BIGF)
[13]  CHECK+(+/I((HF×Q÷+/F)-F0)÷F0)+I(THETA-H×THETA)÷1+FF
[14]  THETA+H×THETA
[15]  F+Q×N+0
[16]  []+'';'13,X2,F10.4,X2,100F7.4' []FMT(ITER,THETA,F0)
[17]  +(1 0 =CHECK(0.0001)/DONE,NEXT3
[18] NEXT3:+(1 0 =ITER(50)/TOP2,END
[19] END:[]+'';[]+'';          CONVERGENCE CRITERION MET'
[20] DONE:[]+'';[]+''
        2
[21]  F100+((((T)0,,F)(1))×(THETA×Q×BIGF)÷(BIGF+THETA×Q×BIGF)×((Q×1GF)+THETA×BIGF)×BIGF+THETA×Q×BIGF
[22]  LAMDA+Q-THETA×Q×V
[23]  TEQF+(THETA×+/(((1)0,,F(1))×[])÷(Q×1GF)×(Q×1GF)+THETA×BIGF)-+/V+(THETA×2)×QV
[24] TOP3:J+J+1
[25]  LAMBDA[J;]+TEMP×[1]
[26]  +(1 0 =F(1)/TOP3,NEXT4
[27] NEXT4:LAMDDAF+(T,1)ρ0
[28]  +(1/(Q+/TIE)=0)/NEXT16,NEXT15
[29] NEXT15:LAMBDDAF[1;]+[1]×(((1+THETA×2)×(THETA×2)-1)×+/Q+/F100)-+/Q+/VXBIGF×2
[30] NEXT16:LAMBDDAF[1;]×(1-1)+(((1+FV[]T1,1)÷V[1,1)×J[];]JV[];]]
        I(1;(1-1)1
```

```
1.31]  LAMBDA=[[4+(T-1)];1+(T-1)]*((LAMBDA+LAMBDA[;T])-(Q[1,1]*LAMBDA[;T])+(T,T)FLAMBDA[
       ;;])]L[(T-1);;(T-1)]
1.32]  SIGMA+((T+1),T+1)F0
1.33]  SIGMA[1;1;T](HLAMBDAF
1.34]  SIGMA[1;1;T+1](SIGMA[+1;1;T](-+/SIGMA[;T;1+(T-1)])
1.35]  SIGMA[T+1;1;1](-+/SIGMA[+1;1]
1.36]  []('';;[]('';'    THE VARIANCE-COVARIANCE MATRIX (SIGMA) IS'
1.37]  []('';;'100F12.6' [FMT(SIGMA)
1.38]  []('';;'THE AVERAGE NUMBER OF REPLICATES IS;   ';L0;(TxT-1)÷2
1.39]  []('';;[]('';'THE VARIANCE-COVARIANCE MATRIX OF THE MLE'S F1,F2,... IS APPROXIMA
       TFLY;
1.40]  []('';;'100F12.6' [FMT(SIGMA÷L0;(TxT-1)÷2)
1.41]  []EXIG;[]('';;[]('';'THE CHI-SQUARE TEST OF THE EQUALITY OF THE "TRUE" TREATMENT F
       ARAMETERS'
1.42]  []('';;'    CHI-SQUARE = ',
1.45]  2x(L0x@0+SH)+(SHx@2)+((L0-SH)x@(SHx(THETAx2)-1)+L0-SH)+(+/((1,T)FB1)x@F0)--+/Q+/BB
       x@(Q;LG0)+THETAxBIGF
1.46]  []('';;'    DEGREES OF FREEDOM = ';T-1
```

APPENDIX B

COMPUTER PROGRAMS THAT FIND
MAXIMUM LIKELIHOOD ESTIMATES OF $\beta_1, \ldots, \beta_s$

This appendix contains two computer programs that find iterative estimates of $\beta_1, \ldots, \beta_s$. The first one, found on pages 174 and 175 is written in the language APL. The second one, found on pages 176 and 177, is written in Fortran. The APL version is discussed first.

The input is the following.

(1) DATA = A $t \times t$ matrix whose $(i,j)^{th}$ element is the number of times T_i is preferred over T_j.

(2) MODEL = A vector of integers defining the model which is to be fit. This is now explained in detail. Suppose there are v independent variables in the experiment. Then the first v elements of MODEL give the order of each variable (e.g. 3 means to include x_i, x_i^2, and x_i^3). The remaining elements are considered in pairs, and correspond to two-way interaction terms which are to be included in the model (e.g. (1,3) occurring after the first v elements means that the term $x_1 x_3$ should be included). Notice that any 3-order or higher order terms can not be included without a modification of the program.

(3) X_o = A $t \times v$ matrix of levels in the experimental design.

The output is the following.

(1) A listing of the vector MODEL.

(2) The t×s matrix X formed by concatenating onto X_o the columns corresponding to the higher order and interaction effects.

(3) The s×t×t matrix where the $(i,j,k)^{th}$ element is $(X)_{ji} - (X)_{ki}$.

(4) The parameter estimates at each iterative step until the convergence criterion is met. The initial estimates are $\hat{\beta}_1 = \cdots = \hat{\beta}_s = 0$. These estimates at each iteration are printed in the same order as the columns of X.

The Newton-Raphson method with a slight modification is used to perform the iteration phase. See Scarborough (1962) for a thorough discussion of the Newton-Raphson method. Springall stated that the iterative estimates of $\theta, \beta_1, \ldots, \beta_s$ converged very slowly. This is also true when only estimating $\beta_1, \ldots, \beta_s$ using the Newton-Raphson method without any modifications. However, it is the author's experience that when the convergence is slow, it is because the estimates are "jumping" from one side of what will be the final estimates to the other. That is, if $\hat{\beta}_1, \ldots, \hat{\beta}_s$ represent the final estimates, then the estimates after the kth iteration may be $\hat{\beta}_1 + \varepsilon_1, \ldots, \hat{\beta}_s + \varepsilon_s$, and the estimates after the $(k+1)^{st}$ iteration may be approximately $\hat{\beta}_1 - \varepsilon_1, \ldots, \hat{\beta}_s - \varepsilon_s$, for some values of ε_i, i=1, ... ,s. Because of this phenomonen, the modification to the Newton-Raphson method is to take the average of the kth and $(k+1)^{st}$ iterative estimate before continuing with the next iteration whenever the direction of convergence changes. For example, if the successive estimates of β_1 are decreasing, and the $(k+1)^{st}$ estimate of β_1 is larger than the kth, then the average of the kth and

$(k+1)^{st}$ estimate replaces the $(k+1)^{st}$ estimate before proceeding to the $(k+2)^{nd}$ iteration. This modification substantially decreased the number of iterations until convergence.

The iteration phase is terminated when the maximum difference between the estimates and the preceding estimates is less than 10^{-6}. If it is desired, this can easily be changed by altering line [55].

As pointed out in Section 2.2, the procedure does not always converge. However, the only time this occurred was when it was attempted to fit a full model, i.e. when s=t-1. When s=t-1, the program in Appendix A can be used to find estimates of $\ln \pi_i$, and then the t equations,

$$\ln \hat{\pi}_i = \sum_{k=1}^{t-1} \hat{\beta}_k x_{ki} , \qquad i=1,\ldots,t,$$

can be solved for $\hat{\beta}_1, \ldots, \hat{\beta}_{t-1}$.

Example. Suppose that

$$DATA = \begin{bmatrix} 0 & 15 & 19 \\ 10 & 0 & 14 \\ 6 & 11 & 0 \end{bmatrix} , \quad X_o = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} , \quad MODEL = [2] .$$

Then an execution of the program yields the following:

$$MODEL = 2 \qquad X = \begin{bmatrix} -1 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$$

Matrix of differences:

$$\begin{bmatrix} 0 & -1 & -2 & \vdots & 0 & 1 & 0 \\ 1 & 0 & -1 & \vdots & -1 & 0 & -1 \\ 2 & 1 & 0 & \vdots & 0 & 1 & 0 \end{bmatrix}$$

Iteration	$\beta_1, \beta_2, \ldots$	
1	-.468182	-.145455
2	-.485958	.319043
3	-.469350	-.144161
AVG	-.477654	.087441
.	.	.
.	.	.
.	.	.
11	-.478947	.084719

As can be seen in the iterative phase in the example, an average is taken between the 3rd and 4th iteration. This occurred because $\hat{\beta}_1$ decreased and $\hat{\beta}_2$ increased from the 1st to the 2nd iteration, while the reverse occurred for the 2nd to the 3rd iteration.

The Fortran program as it currently stands can only be used for fitting a quadratic model. It can, however, be modified to be as general as the APL program. The matrices DATA and X need to be read in. This is done in line numbers 3-8. These statements in addition to the data itself at the end of the program will need to be changed according to the problem at hand.

A peculiar occurrence was made evident by running the Fortran program for the same example as presented above. It turned out that the Fortran program converged to the same final estimates in only 4 iterations instead of 11 iterations. The procedure in each program is the same, but no explanation for the difference in the speed of convergence for the two programs is presently known.

```
      ∇ MODEL ESTIM X;I;J;T;K;MATRIX;XDIFF;S;ITER;P;P1;P2;P3;COR;
        D;DETERM;L;N;BETAOLD;DIFFOLD;DIFF;FLAG;LLL
[1]     BETA←(1,ρX[1;])ρ0
[2]     I←0
[3]     N←ρX[1;]
[4]  L1:→(1 0 =I<N)/L2,L5
[5]  L2:I←I+1
[6]     →(1 0 =MODEL[I]≤1)/L1,L3
[7]  L3:K←1
[8]  L4:K←K+1
[9]     X←X,X[;I]⋆K
[10]    →(1 0 =K<MODEL[I])/L4,L5A
[11] L5A:→(1 0 =I=N)/L5,L2
[12] L5:→(1 0 =N=ρMODEL)/T1,L6
[13] L6:I←I+2
[14]    X←X,X[;MODEL[I-1]]×X[;MODEL[I]]
[15]    →(1 0 =I<(ρMODEL)/L6,T1
[16] T1:'      MODEL = ';MODEL
[17]    □←'';□←'';'         X ='
[18]    □←'';□←'';X
[19]    XDIFF←((S←ρX[1;]),T,(T←ρX[;1]))ρ0
[20]    I←0
[21] T2:I←I+1
[22]    J←I
[23] T3:J←J+1
[24]    XDIFF[;I;J]←X[I;]-X[J;]
[25]    XDIFF[;J;I]←-XDIFF[;I;J]
[26]    →(1 0 =J<T)/T3,T4
[27] T4:→(1 0 =I<((T-1))/T2,T5
[28] T5:□←'';□←'';□←'';'MATRIX OF DIFFERENCES'
[29]    □←'';□←'';XDIFF
[30]    MATRIX←((S,T,T)ρDATA)×XDIFF
[31]    FLAG←ITER←0
[32]    □←'';□←'';□←'';□←'';'ITERATION      B0,B1,B2,...'
[33]    COR←(1,S)ρ0
[34] T6:P←⍋×BETA+.×⍙⊗X
[35]    P1←(T,T)ρ,P
[36]    P2←⍸P1
[37]    P3←P1÷P1+P2
[38]    EQNS←-+/+/MATRIX×(S,T,T)ρP3
[39]    D←(S,S)ρ0
[40]    MAT←(S,S,S)ρ0
[41]    L←0
[42] T7:L←L+1
[43]    D[;L]←+/+/MATRIX×(S,T,T)ρ((T,T)ρX[;L])×P3-(P1×(⍙((T,T)ρX[;L])×P1)+((T,T)ρX[;L])×P1)-(P1+P2)⋆2
[44]    MAT[;;L]←(S.S)ρD[;L]
[45]    MAT[L;;L]←EQNS
```

```
[46]   →(1 0 =L<S)/T7,T8
[47] T8:DETERM←DET D
[48]   I←0
[49] T9:I←I+1
[50]   COR[1;I]←(DET MAT[I;;])÷DETERM
[51]   →(1 0 =I<S)/T9,T10
[52] T10:ITER←ITER+1
[53]   BETAOLD←BETA
[54]   BETA←BETA+COR
[55]   'I5,X5,30F10.6' □FMT((1,S+1)⍴ITER,BETA)
[56]   →(1 0 =(⌈/|BETA-BETAOLD)<1E‾6)/T12,T11
[57] T11:→(1 0 =ITER<50)/T11A,0
[58] T11A:→(1 0 =FLAG=0)/T11B,T11C
[59] T11B:FLAG←1
[60]   →T6
[61] T11C:→(1 0 =FLAG=1)/T11D,T11E
[62] T11D:DIFFOLD←BETA-BETAOLD
[63]   FLAG←2
[64]   →T6
[65] T11E:DIFF←BETA-BETAOLD
[66]   LLL←0
[67] AAA:LLL←LLL+1
[68]   →(1 0 =((×DIFF[1;LLL])=×DIFFOLD[1;LLL]))/BBB,T11F
[69] BBB:→(1 0 =LLL=S)/T11G,AAA
[70] T11F:BETA←(BETA+BETAOLD)÷2
[71]   ' AVG ';'X5,30F10.6' □FMT((1,S)⍴BETA)
[72]   FLAG←1
[73] T11G:DIFFOLD←DIFF
[74]   →T6
[75] T12:'     CONVERGENCE CRITERION MET'
```

```
0001        DIMENSION DATA(5,5),XDIFF(2,5,5),XDIFFM(2,5,5),P(5),PRATIO(5,5),
0002       1X(5,2),BETA(2),D(2,2),MAT(100,3)
0003        READ(5,100)((DATA(I,J),J=1,3),I=1,3)
0004    100 FORMAT(12F2.0)
0005        WRITE(6,101)((DATA(I,J),J=1,3),I=1,3)
0006    101 FORMAT(20X,'DATA'///3(3F6.0//))
0007        READ(5,102)(X(I,1),I=1,3)
0008    102 FORMAT(3F2.0)
0009        DO 1 I=1,3
0010      1 X(I,2)=X(I,1)**2
0011        DO 2 I=1,2
0012        DO 2 J=1,3
0013        DO 2 K=1,3
0014      2 XDIFF(I,J,K)=X(J,I)-X(K,I)
0015        WRITE(6,103)((X(I,J),J=1,2),I=1,3)
0016    103 FORMAT('1',20X,'X'///3(2F6.1//))
0017        WRITE(6,104)(((XDIFF(I,J,K),K=1,3),J=1,3),I=1,2)
0018    104 FORMAT('1',20X,'XDIFF'///3(3F6.0//)///3(3F6.0//))
0019        BETA(1)=0.0
0020        BETA(2)=0.0
0021        CALL ESTIM(DATA,3,100,ITER,X,XDIFF,BETA,D,DET,MAT,M)
0022        WRITE(6,105)ITER,BETA(1),BETA(2)
0023    105 FORMAT('1',10X,'NUMBER OF ITERATIONS'//15X,I6//20X,'BETA'///
0024       12F15.6)
0025        WRITE(6,106)DET,D(1,1),D(1,2),D(2,1),D(2,2),
0026       1((MAT(I,J),J=1,3),I=1,M)
0027    106 FORMAT('1',10X,F15.6//4F20.6////100(/3F20.6))
0028        STOP
0029        END
0030        SUBROUTINE ESTIM(DATA,IIT,NITER,ITER,X,XDIFF,BETA,D,DET,MAT,M)
0031        DIMENSION DATA(5,5),XDIFF(2,5,5),XDIFFM(2,5,5),P(5),D(2,2),
0032       1PRATIO(5,5),X(5,2),BETA(2),BOLD(2),DOLD(2),DIFF(2),MAT(100,3)
0033        DO 1 I=1,2
0034        DO 1 J=1,IIT
0035        DO 1 K=1,IIT
0036      1 XDIFFM(I,J,K)=DATA(J,K)*XDIFF(I,J,K)
0037        M=0
0038        ITER=0
0039        IFLAG=0
0040     10 DO 2 I=1,IIT
0041      2 P(I)=EXP(BETA(1)*X(I,1)+BETA(2)*X(I,2))
0042        M=M+1
0043        MAT(M,1)=FLOAT(ITER)
0044        MAT(M,2)=BETA(1)
0045        MAT(M,3)=BETA(2)
0046        FUNC1=0.0
0047        FUNC2=0.0
0048        DO 3 I=1,IIT
0049        DO 3 J=1,IIT
0050        PRATIO(I,J)=P(J)/(P(I)+P(J))
```

```
0051         FUNC1=FUNC1+XDIFFM(1,I,J)*PRATIO(I,J)
0052       3 FUNC2=FUNC2+XDIFFM(2,I,J)*PRATIO(I,J)
0053         DO 44 IR=1,2
0054         DO 44 L=1,2
0055         DD=0.0
0056         DO 4 I=1,IIT
0057         DO 4 J=1,IIT
0058         IF(I.EQ.J) GO TO 4
0059         DD=DD+XDIFFM(IR,I,J)*(X(J,L)*PRATIO(I,J)-
0060        1(X(I,L)*PRATIO(I,J)*PRATIO(J,I)+X(J,L)*PRATIO(I,J)**2))
0061       4 CONTINUE
0062      44 D(IR,L)=DD
0063         DET=D(1,1)*D(2,2)-D(1,2)*D(2,1)
0064         ITER=ITER+1
0065         BOLD(1)=BETA(1)
0066         BOLD(2)=BETA(2)
0067         BETA(1)=BETA(1)+(FUNC2*D(1,2)-FUNC1*D(2,2))/DET
0068         BETA(2)=BETA(2)+(FUNC1*D(2,1)-FUNC2*D(1,1))/DET
0069         DO 5 I=1,2
0070   .     DIF=ABS(BETA(I)-BOLD(I))
0071         IF(DIF.GT..000001) GO TO 11
0072       5 CONTINUE
0073         RETURN
0074      11 IF(ITER.GE.NITER) RETURN
0075         IF(IFLAG.EQ.2) GO TO 14
0076         IF(IFLAG.EQ.1) GO TO 13
0077         IFLAG=1
0078         GO TO 10
0079      13 DOLD(1)=BETA(1)-BOLD(1)
0080         DOLD(2)=BETA(2)-BOLD(2)
0081         IFLAG=2
0082         GO TO 10
0083      14 DIFF(1)=BETA(1)-BOLD(1)
0084         DIFF(2)=BETA(2)-BOLD(2)
0085         IF(SIGN(DIFF(1),DOLD(1)).NE.DIFF(1)) GO TO 15
0086         IF(SIGN(DIFF(2),DOLD(2)).EQ.DIFF(2)) GO TO 16
0087      15 BETA(1)=(BETA(1)+BOLD(1))/2.
0088         BETA(2)=(BETA(2)+BOLD(2))/2.
0089         M=M+1
0090         MAT(M,1)=-1
0091         MAT(M,2)=BETA(1)
0092         MAT(M,3)=BETA(2)
0093         IFLAG=1
0094      16 DOLD(1)=DIFF(1)
0095         DOLD(2)=DIFF(2)
0096         GO TO 10
0097         END
0098 //GO.SYSIN DD *
0099  0151910 014 611 0
0100 -1 0 1
```

BIBLIOGRAPHY

Abelson, R. M. and Bradley, R. A. (1954). A 2×2 factorial with paired
    comparisons. Biometrics 10, 487-502.

Beaver, R. J. (1977). Weighted least squares response surface fitting
    in factorial paired comparisons. Commun. Statist. A6, 1275-87.

Box, G. E. P. and Lucas, H. L. (1959). Design of experiments in
    nonlinear situations. Biometrika 46, 77-90.

Bradley, R. A. (1953). Some statistical methods in taste testing and
    quality evaluation. Biometrics 9, 22-38.

Bradley, R. A. (1955). Rank analysis of incomplete block designs.
    III. Some large-sample results on estimation and power for a
    method of paired comparisons. Biometrika 42, 450-70.

Bradley, R. A. (1976). Science, statistics, and paired comparisons.
    Biometrics 32, 213-32.

Bradley, R. A. and Terry, M. E. (1952). The rank analysis of incom-
    plete block designs. I. The method of paired comparisons.
    Biometrika 39, 324-45.

David, H. A. (1963). The Method of Paired Comparisons. Griffin,
    London.

Davidson, R. R. (1970). On extending the Bradley-Terry model to
    accommodate ties in paired comparison experiments. J. Amer.
    Statist. Assoc. 65, 317-28.

Davidson, R. R. and Bradley, R. A. (1970). Multivariate paired compar-
    isons: Some large sample results on estimation and tests of
    equality of preference. In Nonparametric Techniques in Statistical
    Inference, (M. L. Puri, ed.). Cambridge Univ. Press, 111-25.

Dykstra, O. (1956). A note on the rank analysis of incomplete block
    designs - applications beyond the scope of existing tables.
    Biometrics 12, 301-6.

Dykstra, O. (1960). Rank analysis of incomplete block designs: a
    method of paired comparisons employing unequal repetitions on
    pairs. Biometrics 16, 176-88.

El-Helbawy, A. T. and Bradley, R. A. (1978). Treatment contrasts in paired comparisons: Large-sample results, applications, and some optimal designs. J. Amer. Statist. Assoc. 73, 831-9.

Ford, L. R., Jr. (1957). Solution of a ranking problem from binary comparisons. Amer. Math. Monthly 64 (8), 28-33.

Graybill, F. A. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth, Belmont, Calif.

Grizzle, J. E., Starmer, C. F. and Koch, G. G. (1969). Analysis of categorical data by linear models. Biometrics 23, 489-504.

Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. (1964). U.S. Department of Commerce, National Bureau of Standards.

Haynam, G. E., Govindarajulu, Z. and Leone, F. C. (1962). Tables of the Cumulative Non-Central Chi-Square Distribution. Case Institute of Technology Statistical Laboratory, Report 104, Cleveland, Ohio.

Kaplan, W. (1952). Advanced Calculus. Addison-Wesley, Reading, Mass.

Larmond, E., Petrasovits, A. and Hill, P. (1968). Applications of multiple paired comparisons in studying the effect of aging and finish on beef tenderness. Canad. J. Anim. Sci. 49, 51-8.

Littell, R. C. and Boyett, J. M. (1977). Designs for R×C factorial paired comparison experiments. Biometrika 64, 73-7.

Quenouille, M. H. and John, J. A. (1971). Paired comparison designs for 2^n factorials. Appl. Statist. 20, 16-24.

Rao, P. V. and Kupper, L. L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. J. Amer. Statist. Assoc. 62, 194-204, Corrigenda 63, 1550.

Scarborough, J. B. (1962). Numerical Mathematical Analysis. The Johns Hopkins Press, Baltimore, Maryland.

Scheffe, H. (1952). An analysis of variance for paired comparisons. J. Amer. Statist. Assoc. 47, 381-400.

Sen, P. K. (1979). Asymptotic properties of maximum likelihood estimators based on conditional specification. Ann. Statist. 7, 1019-33.

Springall, A. (1973). Response surface fitting using a generalization of the Bradley-Terry paired comparison model. Appl. Statist. 22, 59-68.

Thompson, W. A. and Singh, J. (1967). The use of limit theorems in paired comparison model building. Psychometrika 32, 255-64.

Thurstone, L. L. (1927). Psychophysical analysis. Amer. J. Psychol. 38, 368-89.

Wilks, S. S. (1962). Mathematical Statistics. John Wiley & Sons, New York.

BIOGRAPHICAL SKETCH

Walter William Offen was born in Harmony, Minnesota, on November 30, 1953, the only American-born child of German immigrants. Six months later his family moved to the Chicago area where he resided through high school.

He attended Northwestern University and received the Bachelor of Arts degree in June, 1975, with a major in mathematics and a minor in economics. He received the degree of Master of Statistics from the University of Florida in June, 1977. Since that time he has been pursuing the degree of Doctor of Philosophy.

During his time in graduate school at the University of Florida, he worked for the Department of Statistics as an assistant in their Institute of Food and Agricultural Sciences consulting unit. He will begin working for Eli Lilly and Company in September, 1980, as a senior statistician.

Walter Offen is married to the former Susan Elaine Hauser. They will reside in Noblesville, Indiana.
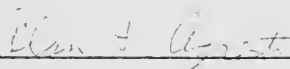
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ramon C. Littell, Chairman
Associate Professor
of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Pejaver V. Rao
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Alan G. Agresti
Associate Professor
of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard F. Matthews
Professor of Food Science
and Human Nutrition

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1980

_____
Dean, Graduate School